On Predictive Density Estimation for Location Families under Integrated Squared Error Loss [1]

Tatsuya Kubokawa[a], Éric Marchand[b], William E. Strawderman[c]
a Department of Economics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, JAPAN (e-mail: tatsuya@e.u-tokyo.ac.jp)
b Université de Sherbrooke, Département de mathématiques, Sherbrooke Qc, CANADA, J1K 2R1 (e-mail: eric.marchand@usherbrooke.ca)
c Rutgers University, Department of Statistics and Biostatistics, 501 Hill Center, Busch Campus, Piscataway, N.J., USA, 08855 (e-mail: straw@stat.rutgers.edu)

Summary

Our investigation concerns the estimation of predictive densities and a study of efficiency as measured by the frequentist risk of such predictive densities with integrated squared error loss. Our findings relate to a $d-$variate spherically symmetric observable $X \sim p_X(\|x-\mu\|^2)$ and the objective of estimating the density of $Y \sim q_Y(\|y-\mu\|^2)$ based on X. We describe Bayes estimation, minimum risk equivariant estimation (MRE), and minimax estimation. We focus on the risk performance of the benchmark minimum risk equivariant estimator, plug-in estimators, and plug-in type estimators with expanded scale. For the multivariate normal case, we make use of a duality result with a point estimation problem bringing into play reflected normal loss. In three or more dimensions (i.e., $d \geq 3$), we show that the MRE predictive density estimator is inadmissible and provide dominating estimators. This brings into play Stein-type results for estimating a multivariate normal mean with a loss which is a concave and increasing function of $\|\hat{\mu}-\mu\|^2$. We also study the phenomenon of improvement on the plug-in density estimator of the form $q_Y(\|y-aX\|^2)\,, 0<a\leq 1\,,$ by a subclass of scale expansions $\frac{1}{c^d}\,q_Y(\|(y-aX)/c\|^2)$ with $c>1$, showing in some cases, inevitably for large enough d, that all choices $c>1$ are dominating estimators. Extensions are obtained for scale mixture of normals including a general inadmissibility result of the MRE estimator for $d \geq 3$.

AMS 2010 subject classifications: 62C20, 62C86, 62F10, 62F15, 62F30
Keywords and phrases: Concave loss, Convolutions, Dominance, Frequentist risk, Inadmissibility, Integrated squared error, Minimax, Minimum risk equivariant, Loss function, Multivariate normal, Predictive density, Restricted parameter space, Scale mixture of normals, Stein estimation.

1 Introduction

1.1 The model and problem

Consider independently distributed

$$X|\mu \sim p(x-\mu), Y|\mu \sim q(y-\mu)\,; x, y, \mu \in \mathbb{R}^d; \tag{1}$$

[1] May 22, 2015

where p and q are known, not necessarily equal, and μ is unknown. For predictive analysis purposes, researchers are interested in specifying a predictive density $\hat{q}(y|x)$ as an estimate of the density $q(y-\mu)$. In turn, such a density may play a surrogate role for generating either future or missing values of Y.
Our interest and motivation here lies in assessing the efficiency of such predictive densities with integrated squared error loss and corresponding frequentist risk, where

$$L(\mu, \hat{q}) = \int_{\mathbb{R}^d} |q(y-\mu) - \hat{q}(y)|^2 \, dy \,,$$
$$R(\mu, \hat{q}) = \int_{\mathbb{R}^d} \{ \int_{\mathbb{R}^d} |q(y-\mu) - \hat{q}(y)|^2 dy \} \, p(x-\mu) \, dx \,. \tag{2}$$

Integrated squared error loss is a familiar distance and is also symmetric in opposition to Kullback-Leibler loss. For normal models, while squared error point estimation loss is dual to Kullback-Leibler loss for prediction, it is rather a bounded loss, namely reflected normal loss, which turns out to dual to integrated squared error loss (see Lemma 3.1). Since bounded losses are appealing to many decision-makers, integrated squared loss is similarly appealing.

The set-up in (1) includes the normal model with

$$X|\mu \sim N_d(\mu, \sigma_X^2 I_d), Y|\mu \sim N_d(\mu, \sigma_Y^2 I_d) \,, \tag{3}$$

scale mixtures of normal distributions (Definition 2.1), and more generally spherically symmetric distributions with

$$X|\mu \sim p_X(\|x-\mu\|^2) \,,\, Y|\mu \sim q_Y(\|y-\mu\|^2) \,, \tag{4}$$

to which the developments of this paper will relate. Such distributions, which include the notable multivariate student case, are interesting alternatives to the multivariate normal in a variety of modelling situations where tails are thicker than those of the normal distribution seem warranted.

1.2 Motivation and overview of findings

As expanded upon below, our findings focus mainly on: **(A)** the risk performance of the benchmark minimum risk equivariant (MRE) estimator $\hat{q}_{mre}$, its inadmissibility quite generally which we establish for $d \geq 3$, and **(B)** improvements on plug-in estimators $q(y - \hat{\mu}(X))$, $y \in \mathbb{R}^d$, where $\hat{\mu}(X)$ is an estimator of μ, obtained by expanding the scale (or variance).

(A) The MRE predictive density estimator is obtained as the generalized Bayes estimator of the density $q(y-\mu)$ with respect to the flat prior $\pi(\mu) = 1$ on $\mathbb{R}^d$. Furthermore, it is minimax and thus represents an important benchmark and an attractive choice as an estimator. These features also hold for Kullback-Leibler (KL) loss (e.g., Liang and Barron, 2004; Kubokawa et al., 2013), defined as $L_{KL}(\mu, \hat{q}) = \int_{\mathbb{R}^d} q(y-\mu) \, \log(\frac{q(y-\mu)}{\hat{q}(y)}) \, dy$. Although $\hat{q}_{mre}$ possesses other interesting features, such as being for $d = 1, 2$ an admissible estimator for normal models (3) under KL loss (Brown, George, Xu, 2008), Komaki (2001) established the inadmissibility of $\hat{q}_{mre}$ for such a normal model, KL loss, for $d \geq 3$, and provided dominating estimators. With such a striking parallel between this result and Stein's inadmissibility of the sample mean as a point estimator of the mean μ of a $N_d(\mu, \sigma_X^2 I_d)$ population under squared error loss, further relationships between Bayesian predictive density estimators and Bayesian point estimators under the same priors were obtained by George, Liang and Xu (2006), Brown, George and Xu (2008), and Fourdrinier et al. (2010) among others.

In Section 2, we provide properties and examples relative to $\hat{q}_{mre}$. A key representation of $\hat{q}_{mre}$; which also applies for KL loss since the MRE estimators coincide; involves a convolution of p and q in (1) (Proposition 2.1, Example 2.2).

In Section 3, for normal models and $d \geq 3$, we establish with Theorem 3.4 the inadmissibility of $\hat{q}_{mre}$ for integrated squared error loss, and provide dominating estimators. We further extend the result to scale mixtures of normals in Section 4.3. These results are achieved by first establishing key relationships between our predictive density estimation problem and a problem of estimating μ based on $X \sim p(x - \mu)$ under a loss of the type $f(\|\hat{\mu} - \mu\|^2)$ where f (which depends on q) is shown to be increasing and concave. Then, we capitalize on known results and/or familiar techniques (e.g., Brandwein and Strawderman, 1991, 1981; Brandwein, Ralescu, Strawderman, 1993) for obtaining dominating point estimators of the usual procedure X, which thus lead to dominating predictive density estimators of $\hat{q}_{mre}$, and the latter's inadmissibility. The dual loss functions that intervene, which include reflected normal loss (Spiring, 1993), are of interest on their own.

Remark 1.1. *Plug-in estimators are ubiquitous in statistical theory and practice. For the univariate normal model (3) and KL loss, Aitchison (1975) showed that the flat prior Bayes procedure (which is $\hat{q}_{mre}$) is a $N(x, \sigma_X^2 + \sigma_Y^2)$ density, and furthermore showed that it dominates the plug-in $N(x, \sigma_Y^2)$ density. Lawless and Fredette (2005) present an instructive approach using a pivotal quantity to obtain KL improvements on plug-in estimators. Fourdrinier et al. (2010) elaborate on plug-in estimators $q(y - \hat{\mu}(X))$ for normal models and KL loss. Their inadmissibility may be directly attributable, in some cases, to the inadmissibility of $\hat{\mu}(X)$ in estimating μ under (a dual) squared error loss (also see part* **B** *below for more on their inefficiency under KL loss). For integrated squared error loss, we do not deal as explicitly as with plug-in estimators of the form $q(y - X)$ since these are invariant and are thus dominated by the MRE estimator $\hat{q}_{mre}$. This explains our focus in* **(A)** *on rather providing dominating predictive density estimators of $\hat{q}_{mre}$.*

(B) Fourdrinier et al. (2011) show, for normal model plug-in estimators $q(y - \hat{\mu}(X))$ and KL loss, that a range of scale expansions always lead to improvements of the form $\hat{q}_c(y; X) = \frac{1}{c^d} q(\frac{y - \hat{\mu}(X)}{c})$ with $c > 1$ regardless of the plugged-in estimator $\hat{\mu}(X)$ and the dimension d as long as it is not degenerate. This may appear paradoxical since the variance associated with the plug-in density $q(y - \hat{\mu}(X))$ matches the variance of the true density $q(y - \mu)$, but it is always best to ignore this true variance and to opt for an estimator $\hat{q}_c$ whose associated variance overestimates the true variance. From the loss function perspective, this is also somewhat paradoxical in that as the estimate $\hat{\mu}(x)$ approaches μ, the loss associated with the plug-in $\hat{q}_1$ approaches 0, while the losses associated with other $\hat{q}_c$'s do not approach 0.

We obtain various findings extending this phenomenon to integrated squared error loss with $\hat{\mu}(X) = aX$: for normal models and $a = 1$ (Section 3.1), normal models and $0 < a < 1$ (Section 3.2), scale mixtures of normal distributions and $a = 1$ (Section 4.2). In Section 3.1, the unbiased predictive density estimator, which is of the form $\hat{q}_c(y; X)$ is also improved on. A surprise arises : in some cases, typically when the dimension d is large enough, all expansions $\hat{q}_c(y; X)$ with $c > 1$ improve on the plug-in estimator $\hat{q}_1(y; X)$! As an example, for normal cases with equal variances, this unusual situation occurs for all $d \geq 4$. Taking c to be infinitely large is of course silly as it becomes equivalent to using a flat density estimate converging to 0, but the integrated squared error penalty is bounded in the normal case (and in some generality), however poor your estimate, and the result brings home another point of view on the inefficiency of the plug-in estimator.

Other findings (Theorems 3.3, 3.5) in this paper relate directly to restricted parameter space settings, where μ belongs to some known subset of $\mathbb{R}^d$, and are derived by exploiting dual relationships between predictive density and point estimation problems as well as restricted parameter space findings (e.g., Marchand and Strawderman, 2004). Although our primary applications and focus relate to predictive density estimation, several of our results also represent point estimation findings under concave loss, complement existing results (see for instance Kubokawa, Marchand and Strawderman, 2015, for further aspects), and are of interest on their own.

The paper is organized as follows. Section 2.1 contains definitions and properties relative to convolutions, scale mixtures of normals, and the integrated squared difference between two multivariate normal densities. Section 2.2 and Section 6 focus on Bayes, best equivariant, and minimax estimation, with properties and accompanying examples. The developments of Section 3 relate to themes **(A), (B)** described above and to the multivariate normal model (3). Section 4 extends several results of Section 3 from multivariate normal to scale mixtures of multivariate normal models, including a $d \geq 3$ inadmissibility result for $\hat{q}_{mre}$ (Section 4.3) and improvements by expansion of scale (Section 4.2).

2 Definitions, preliminary results. Bayes, best equivariant and minimax estimation

2.1 Some definitions and preliminary results

We collect here definitions and properties which will be useful throughout the paper. Distributions in (1) include the subclass of scale mixture of normals, with examples given by the multivariate Cauchy, Student, Logistic, Laplace, Generalized Hyperbolic, Exponential Power distributions, among others (e.g., Andrews and Mallows, 1974).

Definition 2.1. *Model (1) is referred to as a scale mixture of normals model whenever*

$$p(x) = \int_{\mathbb{R}_+} \phi(\frac{x}{v^{1/2}})\, v^{-d/2}\, dG(v)\,,\; q(y) = \int_{\mathbb{R}_+} \phi(\frac{y}{w^{1/2}})\, w^{-d/2}\, dH(w)\,, \tag{5}$$

for $x, y \in \mathbb{R}^d$, where ϕ is (hereafter) taken to be the normal $N_d(0, I_d)$ density, and $V \sim G$, $W \sim H$ are independently distributed mixing random variables on $\mathbb{R}_+$, for which we further assume that $E(V^{-d/2})$ and $E(W^{-d/2})$ are finite. We will denote such models as $X \sim SN_d(G)$ and $Y \sim SN_d(H)$.

Convolutions $p * q$ will be omnipresent in this paper (e.g., Lemma 2.4) and are given by $p * q(t) = \int_{\mathbb{R}^d} p(t-u)\, q(u)\, du$, $t \in \mathbb{R}^d$, for densities p and q. Just as it is the case for the subclass of normal distributions, the above subclass of scale mixture of normals is closed with respect to convolutions.

Lemma 2.1. *For $X \sim SN_d(G)$ and $Y \sim SN_d(H)$, we have $X + Y \sim SN_d(F)$ where F is the cumulative distribution function (cdf) of $T =^d V + W$.*

Proof. Since, conditionally on (V, W), X and Y are independently distributed as $N_d(0, V I_d)$ $N_d(0, W I_d)$ respectively, it follows that $X + Y | V, W \sim N_d(0, (V + W) I_d)$ whence the result. □

The following result, of which the latter part gives the integrated squared difference between multivariate normal densities, will be used several times. A generalization is given below in Lemma 4.1.

Lemma 2.2. *We have for all $\mu_1, \mu_2 \in \mathbb{R}^d$ and $\sigma_1, \sigma_2 \in \mathbb{R}_+$:*

$$\int_{\mathbb{R}^d} \phi(\frac{y-\mu_1}{\sigma_1})\, \phi(\frac{y-\mu_2}{\sigma_2})\, dy \;=\; (\frac{\sigma_1^2\, \sigma_2^2}{\sigma_1^2 + \sigma_2^2})^{d/2}\, \phi(\frac{\mu_1 - \mu_2}{\sqrt{\sigma_1^2 + \sigma_2^2}})\,, \tag{6}$$

$$\int_{\mathbb{R}^d} (\frac{1}{\sigma_1^d}\, \phi(\frac{y-\mu_1}{\sigma_1}) - \frac{1}{\sigma_2^d}\, \phi(\frac{y-\mu_2}{\sigma_2})\,)^2\, dy = \frac{1}{(4\pi\sigma_1^2)^{d/2}} + \frac{1}{(4\pi\sigma_2^2)^{d/2}} - \frac{2}{(\sigma_1^2+\sigma_2^2)^{d/2}}\, \phi(\frac{\mu_1-\mu_2}{\sqrt{\sigma_1^2+\sigma_2^2}})\,. \tag{7}$$

Proof. Identity (6) is readily verified. For (7), expand the square on the left-hand side to obtain

$$\frac{1}{(\sigma_1^2)^d} \int_{\mathbb{R}^d} \phi^2(\frac{y-\mu_1}{\sigma_1})\, dy + \frac{1}{(\sigma_2^2)^d} \int_{\mathbb{R}^d} \phi^2(\frac{y-\mu_2}{\sigma_2})\, dy - \frac{2}{(\sigma_1\sigma_2)^d} \int_{\mathbb{R}^d} \phi(\frac{y-\mu_1}{\sigma_1})\, \phi(\frac{y-\mu_2}{\sigma_2})\, dy\,.$$

Applying identity (6) to these three terms leads to (7). □

2.2 Bayes and minimum risk equivariant estimators

As in the case of Kullback-Leibler loss, Bayes estimators under integrated squared error loss are simply given by the predictive density $q(y|x)$.

Lemma 2.3. *For model (1), integrated squared error loss, a prior density π for μ, and a posterior density $\pi(\mu|x)$ with respect to measure ν, the Bayes predictive density estimator of $q(y-\mu)$, $y \in \mathbb{R}^d$, is given by*

$$\hat{q}_\pi(y;x) = \int_{\mathbb{R}^d} q(y-\mu)\, \pi(\mu|x)\, d\nu(\mu)\,. \tag{8}$$

Proof. The expected posterior loss for estimator $\hat{q}(\cdot)$ is given by

$$\int_{\mathbb{R}^d} \{ \int_{\mathbb{R}^d} (q(y-\mu) - \hat{q}(y))^2 dy \}\, \pi(\mu|x)\, d\nu(\mu)\,.$$

Interchanging the order of integration, we see that for each y the minimizing $\hat{q}(y)$ is the posterior expectation $E^{\mu|x}(q(y-\mu))$ which, being a density as a function of y, yields the result. □

For location models as in (1) with spherically symmetric q, we obtain an interesting representation when the family of posterior densities is a location family.

Lemma 2.4. *In Lemma 2.3, whenever the prior density is supported on $\mathbb{R}^d$ and the posterior density is of the form $\pi(\mu|x) = g(\mu - \hat{\mu}(x))$, the Bayes predictive density estimator of $q(y-\mu)$, $y \in \mathbb{R}^d$, is equal to $q * g(y - \hat{\mu}(x))$, where $q * g$ is the convolution of q and g.*

Proof. From (8) and with $\pi(\mu|x) = g(\mu - \hat{\mu}(x))$, the Bayes predictive density of $q(y-\mu)$ is equal to $\int_{\mathbb{R}^d} q(y-\mu)\, g(\mu - \hat{\mu}(x))\, d\mu \; = \int_{\mathbb{R}^d} q(y - \hat{\mu}(x) - \mu')\, g(\mu')\, d\mu' = (q*g)(y - \hat{\mu}(x))$. □

Remark 2.1. *Since the Bayes predictive density estimators coincide for Kullback-Leibler and integrated squared error losses, the above lemma and the examples that follow apply as well to Kullback-Leibler loss.*

Example 2.1. *Consider the normal case (3) with $\pi(\mu) \sim N_d(\theta, \tau^2 I_d)$. Since $\mu|x \sim N_d(\hat{\mu}(x), (\tau')^2 I_d)$ with $\hat{\mu}(x) = \frac{\tau^2 x}{\sigma_X^2 + \tau^2} + \frac{\sigma_X^2 \theta}{\sigma_X^2 + \tau^2}$ and $(\tau')^2 = \frac{\sigma_X^2 \tau^2}{\sigma_X^2 + \tau^2}$, the conditions of Lemma 2.4 are satisfied with $q \sim N_d(0, \sigma_Y^2 I_d)$, $g \sim N_d(0, (\tau')^2 I_d)$, and we obtain $\hat{q}_\pi(y;x) \sim N_d(\hat{\mu}(x), (\sigma_Y^2 + (\tau')^2) I_d)$. We further point out, as deduced from above, that all predictive densities $N_d(aX + b, (\sigma_Y^2 + a\sigma_X^2) I_d)$ with $0 \leq a < 1$, $b \in \mathbb{R}$, are unique Bayes estimators with finite Bayes risks and hence admissible. On the other hand, we will show in Section 3 that the MRE estimator (i.e., $a = 1, b = 0$) is inadmissible for $d \geq 3$.*

Example 2.2. *Consider model (1) with the uniform prior $\pi(\mu) = 1$ on $\mathbb{R}^d$ and with the corresponding Bayes predictive density estimator coinciding with the MRE estimator (see the next paragraph). This gives us: $x - \mu|x \sim p$ and Lemma 2.4 applies with $g(y) = \bar{p}(y) = p(-y)$, yielding the representation $\hat{q}_{mre}(y;x) = (q * \bar{p})(y - x)$. Moreover, if p is spherically symmetric as in (4), we obtain $\hat{q}_{mre}(y;x) = (q * p)(y - x)$.*

The MRE predictive density estimator can be derived as the Bayes rule with respect to the Haar invariant prior $\pi(\mu) = 1$ for μ, and is minimax. This follows as the problem is invariant under the group of location changes (including the choice of loss), and from a general representation for the minimum risk equivariant estimator as the Bayes estimator associated with the corresponding Haar measure (e.g., Eaton, 1989), and with the minimaxity following from Kiefer (1959). The following Proposition summarizes the above and provides a direct, instructive and alternative approach in deriving the minimum risk equivariant predictive density estimator under integrated squared error loss, which is analogous to results obtained by Murray (1977) or Kubokawa et al. (2013) for Kullback-Leibler loss.

Proposition 2.1. *The minimum risk equivariant estimator of* $q(y-\mu)$, $y \in \mathbb{R}^d$, *for model (1) and integrated squared error loss is given by*

$$\hat{q}_{mre}(y;x) \;=\; q * \bar{p}(y-x)\,, \tag{9}$$

with $\bar{p}(t) = p(-t)$ *for all* t, *and matches the Bayes predictive density with respect to the uniform prior on* $\mathbb{R}^d$ *given in Example 2.3* **(a)**. *Furthermore,* $\hat{q}_{mre}(\cdot;X)$ *is a minimax predictive density estimator.*

Proof. While Kiefer's result, mentioned above, gives minimaxity quite generally for the MRE estimator, minimaxity is established directly in Section 6 for the general location case via the argument of Girschick and Savage (1951), using a (least favourable) sequence of Uniform priors on the product sets $\{\mu : |\mu_i| < k/2\,, i = 1, \ldots, d\}$, $k = 1, 2, \ldots$. We give a similar direct, but simpler argument specifically for the normal case in Section 2.3.

For the minimum risk equivariance property, we only need to establish (9). First, equivariant estimators under the additive group of transformation $(x, y) \to (x + a, y + a)$; $a \in \mathbb{R}^d$; satisfy the identity

$$\hat{q}(y;x) = \hat{q}(y-x;0) \text{ for all } x, y \in \mathbb{R}^d\,, \tag{10}$$

as seen by setting $a = -x$. The risk of such estimators is constant in $\mu \in \mathbb{R}^d$ and given by

$$\begin{aligned} R(\mu,\hat{q}) &= \int_{\mathbb{R}^d}\{\int_{\mathbb{R}^d}(q(y-\mu)-\hat{q}(y-x;0))^2 dy\}\; p(x-\mu)dx \\ &= \int_{\mathbb{R}^d}\{\int_{\mathbb{R}^d}(q(y)-\hat{q}(y-x;0))^2 dy\}\; p(x)dx \\ &= \int_{\mathbb{R}^d}\{\int_{\mathbb{R}^d}(q(u+v)-\hat{q}(v;0))^2\, p(u)\, du\}\; dv\,, \end{aligned}$$

with the last equality obtained with transformation $(x, y) \to (u = x, v = y - x)$. Now, for all $v \in \mathbb{R}^d$, the inner integral above is minimized by choosing $\hat{q}(v; 0)$ to be the expected value of $q(v + U)$ with $U \sim p$, i.e. $\hat{q}_{\text{opt}}(v) = \int_{\mathbb{R}^d} q(u + v)\, p(u)\, du$. Finally, this along with (10) tell us that

$$\hat{q}_{\text{mre}}(y|\,x) = \hat{q}_{\text{opt}}(y-x) = \int_{\mathbb{R}^d} q(u+y-x)\, p(u)\, du \;=\; \int_{\mathbb{R}^d} q(y-x-u)\, p(-u)\, du \;=\; q*\bar{p}(y-x)\,. \quad \square$$

We conclude this section with illustrative evaluations of $\hat{q}_{\text{mre}}$.

Example 2.3. *Consider scale mixtures of normals densities $X - \mu \sim SN_d(G)$, $Y - \mu \sim SN_d(H)$ as in Definition 2.1. It follows as a particular case of Example 2.2 and from Lemma 2.1 that*

$$\hat{q}_{mre}(y;x) = \int_{\mathbb{R}_+} \phi(\frac{y-x}{t^{1/2}})\, t^{-d/2}\, dF(t)\,, \tag{11}$$

with F the cdf of $T =^d W + V$. For the normal case (3) with $X - \mu \sim N_d(0, \sigma_Y^2 I_d)$, $Y - \mu \sim N_d(0, \sigma_X^2 I_d)$ (i.e., the degenerate case of a scale mixtures of normals), we obtain immediately that $\hat{q}_{mre}(y;x) = N_d(x, (\sigma_X^2 + \sigma_Y^2) I_d)$.
A prominent scale mixture of normals example is the multivariate Student $T(\nu, \sigma)$ with degrees of freedom $\nu > 0$ and scale parameter $\sigma > 0$. In (1), this corresponds to $X - \mu \sim p \sim T(\nu_1, \sigma_1)$, $Y - \mu \sim q \sim T(\nu_2, \sigma_2)$ where the density of $T(\nu, \sigma)$ is given by

$$\frac{\Gamma(\frac{1}{2}(d+\nu))}{(\pi \nu \sigma^2)^{d/2}\; \Gamma(\frac{\nu}{2})}\; (1 + \frac{\|t\|^2}{\nu\, \sigma^2})^{-\frac{1}{2}(d+\nu))}\,.$$

Example 2.2 tells us that $\hat{q}_{mre}(y;x) = (q * p)(y - x)$. Such a convolution density, including cases where one of the densities is that of a normal distribution, has arisen in other settings and been analyzed by others (e.g., Nason 2006; Berg and Vignat, 2010). The particular case of a multivariate Cauchy ($\nu_1 = \nu_2 = 1$) gives rise, simply, to

$$\hat{q}_{mre}(y;x) \;=\; \frac{\Gamma(\frac{1}{2}(d+1))}{\pi^{\frac{d+1}{2}}\; (\sigma_1 + \sigma_2)^d}\; (1 + \frac{\|y-x\|^2}{(\sigma_1 + \sigma_2)^2})^{-\frac{1}{2}(d+1)}$$

since $T(1, \sigma_1) * T(1, \sigma_2) = T(1, \sigma_1 + \sigma_2)$.

3 Plug-in type estimators: the normal case

3.1 Duality and the efficiency of density estimators $N_d(\hat{\mu}(X), c^2 \sigma_Y^2\, I_d)$

We consider here normal model (3) and the performance of density estimators $\hat{q}_{c^2,\hat{\mu}} \sim N_d(\hat{\mu}(X), c^2 \sigma_Y^2\, I_d)$, which combine both a plug-in component with $\hat{\mu}(X)$ being an estimate of μ, and a modification of variance component for $c^2 \neq 1$. As for Kullback-Leibler loss (Fourdrinier et al. 2011), we demonstrate that the efficiency of such estimators relates to: **(i)** the efficiency of the point estimator $\hat{\mu}(X)$ in estimating μ, as well as **(ii)** the degree of variance expansion governed by the choice of $c^2 > 1$. With respect to **(i)** and the duality with the point estimation problem, it is a reflected normal loss that arises, which we denote and define as

$$L_\gamma(\mu, \hat{\mu}) \;=\; 1 - e^{-\frac{\|\hat{\mu}(x) - \mu\|^2}{2\gamma}}\;, \text{ with } \gamma > 0\,, \tag{12}$$

in contrast to squared-error loss which intervenes in duality for Kullback-Leibler loss.

Lemma 3.1. *For estimating a multivariate normal density of $Y \sim N_d(\mu, \sigma_Y^2 I_d)$, the integrated squared error loss of the density estimate $\hat{q}_{c^2,\hat{\mu}} \sim N_d(\hat{\mu}(x), c^2 \sigma_Y^2 I_d)$ is given by*

$$\frac{1}{(\sigma_Y^2)^{d/2}} \left(\frac{1}{(4\pi)^{d/2}} + \frac{1}{(4\pi c^2)^{d/2}} - \frac{2}{(2\pi(c^2+1))^{d/2}}\, e^{-\frac{(\|\hat{\mu}(x) - \mu\|)^2}{2\sigma_Y^2 (c^2+1)}} \right), \tag{13}$$

Proof. This is a direct application of (7) with $\mu_1 = \mu, \mu_2 = \hat{\mu}(x)$, $\sigma_1^2 = \sigma_Y^2$, and $\sigma_2^2 = c^2\sigma_Y^2$. $\square$

Theorem 3.1. **(a)** *For fixed c^2, the frequentist risk of the density estimator $\hat{q}_{c^2,\hat{\mu}} \sim N_d(\hat{\mu}(x), c^2\sigma_Y^2 I_d)$ under integrated squared error loss is equal to the frequentist risk of the point estimator $\hat{\mu}(X)$ of μ under loss $a + b\, L_{\gamma_0}(\mu, \hat{\mu})$, with $a+b = \frac{1+c^d}{(4\pi c^2\sigma_Y^2)^{d/2}}, b = \frac{2}{(2\pi(c^2+1)\sigma_Y^2)^{d/2}}$, and $\gamma_0 = (c^2+1)\,\sigma_Y^2$. Namely, $\hat{q}_{c^2,\hat{\mu}_1} \sim N_d(\hat{\mu}_1(X), c^2\,\sigma_Y^2 I_d)$ improves on $\hat{q}_{c^2,\hat{\mu}_2} \sim N_d(\hat{\mu}_2(X), c^2\,\sigma_Y^2 I_d)$ iff $\hat{\mu}_1(X)$ improves on $\hat{\mu}_2(X)$ under reflected normal loss $L_{\gamma_0}(\mu, \hat{\mu})$.*

(b) *For $\hat{\mu}(X) = X$, the risk $R(\mu, \hat{q}_{c^2,\hat{\mu}})$ is constant as a function of μ, and given by*

$$\frac{1}{(2\pi\sigma_Y^2)^{d/2}} \left((\frac{1}{2})^{d/2} + \frac{1}{(2c^2)^{d/2}} - \frac{2}{(r+c^2+1)^{d/2}} \right), \tag{14}$$

with $r = \frac{\sigma_X^2}{\sigma_Y^2}$. For all d, the constant (and minimax) risk of $\hat{q}_{mre}$, corresponding to the optimal choice $c^2 = 1+r$ is equal to

$$R(\mu, \hat{q}_{mre}) = \frac{1}{(4\pi\sigma_Y^2)^{d/2}} - \frac{1}{(4\pi(\sigma_X^2+\sigma_Y^2))^{d/2}}. \tag{15}$$

(c) *Furthermore, all estimators $\hat{q}_{c^2,\hat{\mu}}$ with $c^2 > 1$ dominate the plug-in or mle estimator $\hat{q}_{1,\hat{\mu}}$ whenever $d \geq d_0 = \frac{\log 4}{\log(1+r/2)}$, and, otherwise for $d < d_0$, the estimator $\hat{q}_{c^2,\hat{\mu}}$ dominates $\hat{q}_{1,\hat{\mu}}$ iff $1 < c^2 < k(d,r)$, where $k(d,r)$ is the unique solution on $(1,\infty)$ in c^2 of the equation*

$$(\frac{1}{2})^{d/2} + 2(\frac{1}{r+c^2+1})^{d/2} - 2(\frac{1}{r+2})^{d/2} - (\frac{1}{2c^2})^{d/2} = 0. \tag{16}$$

Proof. Part **(a)** follows directly from Lemma 3.1. For part **(b)**, use (13) with $\hat{\mu}(x) = x$ and $\hat{q}_{c^2} \sim N_d(x, c^2\, I_d)$ to obtain

$$R(\mu, \hat{q}_{c^2}) = \frac{1}{(\sigma_y^2)^{d/2}} \left(\frac{1}{(4\pi)^{d/2}} + \frac{1}{(4\pi c^2)^{d/2}} - \frac{2}{(c^2+1)^{d/2}}\, E_\mu \left(\phi(\frac{X-\mu}{\sqrt{\sigma_Y^2(c^2+1)}}) \right) \right).$$

The use of identity (6) leads to (14). Now, set $\psi(c^2) = \frac{1}{(2c^2)^{d/2}} - \frac{2}{(r+c^2+1)^{d/2}}$ so that $(2\pi\sigma_y^2)^{d/2}\, R(\mu, \hat{q}_{c^2}) = (\frac{1}{2})^{d/2} + \psi(c^2)$, and observe that $\text{sgn}(\psi'(c^2)) = \text{sgn}\{(2c^2)^{1+d/2} - (r + c^2+1)^{1+d/2}\}$ for $c > 0$. From this, we infer that $\psi'(c^2)$ changes signs once, on $(0,\infty)$, from $-$ to $+$ at $c^2 = 1+r$, which along with the evaluation of (14) for $c^2 = 1+r$ establishes part **(b)**. For part **(c)**, the comparison of $\hat{q}_{c^2}$ with $\hat{q}_1$ for $c^2 > 1$ hedges on the sign of $\psi(c^2)-\psi(1)$ for $c > 1$. From above, we know that $\psi(c^2) - \psi(1)$ is either negative for all $c^2 > 1$, or negative iff $1 < c^2 < k(d,r)$. Finally, we have $\lim_{c^2\to\infty}\{\psi(c^2) - \psi(1)\} = \frac{2}{(2+r)^{d/2}} - \frac{1}{2^{d/2}} \leq 0$ if and only if $d \geq d_0$, concluding the proof. $\square$

Example 3.1. *For equal variances (i.e., $r = 1$) σ_X^2 and σ_Y^2, we obtain $d_0 \approx 3.419$ so that universal dominance for all choices $\hat{q}_{c^2,\hat{\mu}}$ with $c^2 > 1$ over $\hat{q}_{1,\hat{\mu}}$ arises for $d \geq 4$. And* *for $d \geq 3$, the cut-off points $k(1,d)$ are given by $k(1,2) = 6$ (exact), $k(1,1) \approx 4.65$, and $k(1,3) \approx 11.47$. We remark upon the fact that d_0 decreases as the ratio $r = \frac{\sigma_X^2}{\sigma_Y^2}$ increases so that the above universal dominance occurs also for (at least) all $p \geq 4$ whenever $\sigma_X^2 > \sigma_Y^2$. For instance if $\sigma_X^2 = 2\sigma_Y^2$ (i.e., $r = 2$), we obtain $d_0 = 2$.*

Remark 3.1. *An analysis of (14) tells us that the ratio of risks between the minimum risk equivariant estimator $\hat{q}_{1+r,X} \sim N_d(X, (\sigma_X^2 + \sigma_Y^2)\, I_d)$ and the plug-in estimator $\hat{q}_{1,X} \sim N_d(X, \sigma_Y^2 I_d)$ is given by $\frac{R(\mu,\hat{q}_{1,X})}{R(\mu,\hat{q}_{1+r,X})} = 2\left(\frac{1-(\frac{1}{1+r/2})^{d/2}}{1-(\frac{1}{1+r})^{d/2}}\right)$. It is easy to verify that this ratio increases in both r and d, and approaches 2 when either r or d increase to ∞. The monotonicity in the ratio of variances $r = \frac{\sigma_X^2}{\sigma_Y^2}$ translates, understandably, to worsening performance of $\hat{q}_1$ as the relative variability of the observable X increases.*

Remark 3.2. *(On the unbiased predictive density estimator) An unbiased predictive density estimator of the density of $Y|\mu$, (i.e., of $\frac{1}{(\sigma_Y^2)^{d/2}}\,\phi(\frac{y-\mu}{\sigma_Y})$) exists whenever $\sigma_X^2 < \sigma_Y^2$ (for the univariate case, see for instance Lehmann and Casella, 1998; or Shao, 1999). Indeed, considering density estimates $q_{c^2,X} \sim N_d(X, c^2\sigma_Y^2 I_d)$, we have from (6)*

$$\int_{\mathbb{R}^d} \frac{1}{(c^2\sigma_Y^2)^{d/2}}\,\phi(\frac{y-x}{c\,\sigma_Y})\,\frac{1}{(\sigma_X^2)^{d/2}}\,\phi(\frac{x-\mu}{\sigma_X})\,dx = \frac{1}{(\sigma_X^2 + c^2\sigma_Y^2)^{d/2}}\,\phi(\frac{y-\mu}{\sqrt{(\sigma_X^2 + c^2\sigma_Y^2)}})\,,$$

so that a $N_d(X, c^2\sigma_Y^2 I_d)$ density is an unbiased estimator of a $N_d(\mu, \sigma_X^2 + c^2\sigma_Y^2 I_d)$ density (pointwise and globally), and the choice $c^2 = 1 - \frac{\sigma_X^2}{\sigma_Y^2}(> 0)$ yields an unbiased estimator of the density of $Y|\mu$. Since X is a complete sufficient statistic, it follows that this estimator is the sole unbiased estimator. [2] *Here, the unbiased predictive density estimator shrinks the variance, instead of expanding it. It will thus, with its risk given by (14) and as already analysed as a function of c^2, perform even worse than the plug-in $\hat{q}_{1,X}$. In fact, it is dominated by the plug-in, the best equivariant estimator, a range of choices $\hat{q}_{c,X}$, $1 - r < c^2 < k_0(d,r)$, and with $k_0(d,r) = +\infty$ as soon as $d \geq -\frac{\log 4}{\log(1-r)}$.*

3.2 Plug-in estimators with $\hat{\mu}(X) = aX$: improvements by expanding the scale

The scale expansion results of the previous section apply to plug-in estimates $\hat{\mu}(x) = x$, and it is natural to investigate whether similar phenomena occur for other plug-in estimates. We thus consider here the performance of estimators $\hat{q}_{c^2,\hat{\mu}} \sim N_d(\hat{\mu}(X), c^2\sigma_Y^2 I_d)$, and with more development for the affine linear case $\hat{\mu}(x) = ax$, $0 < a \leq 1$. As seen in Section 3.1, there exists for $a = 1$ an optimal choice (i.e., $c^2 = 1 + r$ with $r = \frac{\sigma_X^2}{\sigma_Y^2}$) of the expansion factor c^2 and, for $d \geq d_0 = \frac{\log 4}{\log(1+r/2)}$, we can expand the variance as much as desired and still dominate the plug-in $N_d(x, \sigma_Y^2 I_d)$. The objective here is to assess whether such results hold for other choices of $\hat{\mu}(X)$ and more specifically: **(i)** to determine a range of variance expansions or values c^2 that lead to improvement, and **(ii)**

[2] The more standard set-up, perhaps, has $\sigma_X^2 = \frac{\sigma_Y^2}{n}$, where n is the size of a sample drawn from X.

to determine whether there exists a universal dominance result for sufficiently large d (i.e., for all $c^2 > 1$). Explicit findings with respect to **(i)** and Kullback-Leibler loss were obtained by Fourdrinier et al. (2011) with the maximum amount of allowable expansion to retain improvement for all μ an increasing function of the infimum squared error risk.

From (13), we start off the risk expression

$$R(\mu, \hat{q}_{c^2,\hat{\mu}}) = \frac{1}{(4\pi\sigma_Y^2)^{d/2}} + \frac{1}{(2\pi\sigma_Y^2)^{d/2}} \left(\frac{1}{(2\,c^2)^{d/2}} - \frac{2}{(c^2+1)^{d/2}}\, E_\mu (e^{-\frac{\|\hat{\mu}(X)-\mu\|^2}{2\sigma_Y^2(c^2+1)}}) \right), \quad (17)$$

and the derivative

$$\frac{\partial}{\partial c^2} R(\mu, \hat{q}_{c^2,\hat{\mu}}) = \frac{1}{(2\pi\sigma_Y^2)^{d/2}} \left(\frac{1}{(c^2+1)^{d/2+1}}\, E_\mu \left(e^{-\frac{Z}{2(c^2+1)}} (d - \frac{Z}{c^2+1}) \right) - \frac{d}{(2c^2)^{d/2+1}} \right), \quad (18)$$

with $Z = \frac{\|\hat{\mu}(X)-\mu\|^2}{\sigma_Y^2}$.

Remark 3.3. *Plug-in estimators $\hat{q}_{1,\hat{\mu}} \sim N_d(\hat{\mu}(X), \sigma_Y^2 I_d)$ with non-degenerate $\hat{\mu}(X)$ can always be improved locally at μ by an expansion $\hat{q}_{c^2,\hat{\mu}} \sim N_d(\hat{\mu}(X), c^2\sigma_Y^2 I_d)$ with $c^2 \in (1, c_0^2(\mu))$ for some $c_0^2(\mu) > 1$ (dependent on $\hat{\mu}(X)$). If possible, global dominance $\mu \in \Theta$ is thus achieved by selecting $c^2 \in (1, \inf_{\mu\in\Theta} c_0^2(\mu)]$. This can be seen by a continuity argument and (18), since*

$$\frac{\partial}{\partial c^2} R(\mu, \hat{q}_{c^2,\hat{\mu}})|_{c^2=1} = \frac{1}{2}\, \frac{1}{(4\pi\sigma_Y^2)^{d/2}} \left(E_\mu (e^{-\frac{Z}{4}} (d - \frac{Z}{2}) - d \right) < 0\,,$$

as $e^{-\frac{y}{4}}(d - \frac{y}{2}) - d \leq 0$ for all $y \geq 0$, with equality iff $y = 0$.

For the particular case $\hat{\mu}(X) = aX, 0 < a < 1$, we arrive at more explicit expressions for the risk and its derivative in (17) and (18) by using the exact distributional result $Z = \frac{\|aX-\mu\|^2}{\sigma_Y^2} \sim a^2\, r\, \chi_d^2(\frac{(a-1)^2\|\mu\|^2}{a^2\sigma_X^2})$, and the mixture representation:

$$Z|L \sim Gamma(\frac{d}{2} + L, 2a^2\, r)\,,\ L \sim \text{Poisson}(\frac{\delta}{2}),\ \text{with } \delta = \frac{(a-1)^2\|\mu\|^2}{a^2\sigma_X^2},\ r = \frac{\sigma_X^2}{\sigma_Y^2}\,.$$

Lemma 3.2. *For $Z = \frac{\|aX-\mu\|^2}{\sigma_Y^2}$, we have*

$$E_\mu \left(e^{-\frac{Z}{2(c^2+1)}} (d - \frac{Z}{c^2+1}) \right) = (d-h)\, \theta^{\frac{d}{2}+1}\, e^{-\frac{h}{2}}\,, \quad (19)$$

with $h = \frac{(a-1)^2\,\|\mu\|^2}{a^2\sigma_X^2 + (c^2+1)\sigma_Y^2}$ and $\theta = \frac{c^2+1}{a^2r+c^2+1}$.

Proof. A calculation yields

$$E_\mu \left(e^{-\frac{Z}{2(c^2+1)}} (d - \frac{Z}{c^2+1})|L \right) = d\,\theta^{\frac{d}{2}+L+1} + 2L\,\theta^{\frac{d}{2}+L}\,(\theta - 1)\,,$$

with $L \sim \text{Poisson}(\delta/2)$. The result follows by the Poisson related evaluations $E(\theta^L) = e^{-h/2}$ and $E(L\,\theta^L) = \frac{\theta\delta}{2}\, e^{-h/2}$, and by collecting terms. $\square$

Setting $\psi_a(c^2) = \frac{\partial}{\partial c^2} R(\mu, \hat{q}_{c^2,\hat{\mu}})$ for $\hat{\mu}(X) = aX$, it thus follows from the above expression and (18) that

$$\psi_a(c^2) \;=\; \frac{1}{(2\pi\sigma_Y^2)^{d/2}}\,\frac{1}{(a^2r + c^2 + 1)^{d/2+1}}\,\left((d-h)\,e^{-h/2} \;-\; d\,(\frac{a^2r + c^2 + 1}{2c^2})^{d/2+1}\right), \quad (20)$$

with h as in Lemma 3.2. Here is now the main result of this subsection.

Theorem 3.2. *For $\hat{\mu}(X) = aX$, $0 < a < 1$, $\hat{q}_{c^2,\hat{\mu}}$ dominates $\hat{q}_{1,\hat{\mu}}$ under integrated squared error loss if and only if $1 < c^2 \le k_a(d)$, where $k_a(d) = \infty$ whenever $d \ge d_0(a) = \frac{\log(4)}{\log(1+\frac{a^2r}{2})}$, and otherwise when $d < d_0(a)$, $k_a(d)$ is the unique solution in $c^2 \in (1,\infty)$ of the equation*

$$(\frac{1}{2})^{d/2} + 2(\frac{1}{c^2 + a^2r + 1})^{d/2} - 2(\frac{1}{a^2r + 2})^{d/2} - (\frac{1}{2c^2})^{d/2} \;=\; 0\,. \quad (21)$$

Remark 3.4. *For $a = 1$, we recover part* **(c)** *of Theorem 3.1 and, namely, the universal in c^2 dominance for $d \ge d_0(1)$. Observe that the universal dominance property for all $c^2 > 1$ is inevitable for large enough d. This is not necessarily the case for other estimators (see footnote 3).*

Proof of Theorem 3.2.

(A) We first prove that, for all $a \in (0,1)$, $\mu \in \mathbb{R}^d$, $\psi_a(c^2)$ changes signs once from $-$ to $+$ as c^2 increases on $[1,\infty)$. We have already established (Remark 3.3) that $\psi_a(1) < 0$. As well, $\psi_a(c^2)$ is clearly negative for $h \ge d$, i.e., $1 \le c^2 \le k_0$, with $k_0 = \left(1 \wedge (\frac{(a-1)^2\,\|\mu\|^2}{d\sigma_Y^2} - a^2r - 1))\right)$. Otherwise for $c^2 > k_0$, it is easy to see that $\psi_a(c^2)$ is increasing in c^2 for $c^2 > k_0$. Finally, the stated property follows since $\lim_{c^2\to\infty}((a^2 + c^2 + 1)^{d/2+1})(2\pi\sigma_Y^2)^{d/2}\,\psi_a(c^2) \;=\; d(1 - (\frac{1}{2})^{d/2+1}) > 0$.

(B) Denote the difference in risks

$$\Delta_{c^2}(\mu) = R(\mu, \hat{q}_{1,\hat{\mu}}) - R(\mu, \hat{q}_{c^2,\hat{\mu}})\,. \quad (22)$$

Locally at μ, it follows from part (A) and for $\hat{\mu}(X) = aX$ that

$$\Delta_{c^2}(\mu) \ge 0 \;\text{ if and only if } 1 < c^2 \le c_0^2(\mu)\,, \quad (23)$$

with equality only at $c^2 = c_0^2(\mu)$.

(C) It is possible that $c_0^2(\mu) = +\infty$ for all $\mu \in \mathbb{R}^d$, not only for $\hat{\mu}(X) = aX$, but also more generally for other $\hat{\mu}$'s. A necessary and general condition for this type of universal dominance to occur, for non-degenerate $\hat{\mu}(X)$ and $T = \frac{\|\hat{\mu}(X)-\mu\|^2}{\sigma_Y^2}$, is

$$R(\mu, \hat{q}_{1,\hat{\mu}}) - \lim_{c\to\infty} R(\mu, \hat{q}_{c^2,\hat{\mu}}) \geq 0 \text{ for all } \mu \iff \sup_{\mu} E_{\mu}(e^{-T/4}) \leq \frac{1}{2}.^{3} \qquad (24)$$

This is a verified from (17) where we obtain $(4\pi\sigma_Y^2)^{d/2} \lim_{c^2\to\infty} R(\mu, \hat{q}_{c^2,\hat{\mu}}) = 1$, and $(4\pi\sigma_Y^2)^{d/2}\, R(\mu, \hat{q}_{1,\hat{\mu}}) = 2 - 2\, E_{\mu}(e^{-T/4})$.

(D) Applying condition (24) to $\hat{\mu}(X) = aX$ and making use of the stochastically increasing property of the family of distributions of $Z = \frac{\|aX-\mu\|^2}{\sigma_Y^2} \sim a^2 r \chi_d^2(\frac{(a-1)^2\|\mu\|^2}{a^2\sigma_X^2})$, with $\|\mu\|^2$ viewed as the parameter, we infer that

$$\sup_{\mu\in\mathbb{R}^d} E_{\mu}(e^{-Z/4}) = E_0(e^{-Z/4}) \;=\; (1+\frac{a^2 r}{2})^{-d/2}.$$

Therefore, condition (24) becomes equivalent to $(1+\frac{a^2r}{2})^{-d/2} \leq \frac{1}{2}$ which is $d \geq d_0(a)$.

(E) We conclude the proof by showing that, for fixed $c^2 > 1$,

$$\Delta_{c^2}(\mu) \geq 0 \text{ for all } \mu \in \mathbb{R}^d \text{ iff } \Delta_{c^2}(0) \geq 0\,. \qquad (25)$$

With the condition $\Delta_{c^2}(0) \geq 0$ equivalent to $1 < c^2 \leq k_a(d)$, as can be seen by making use of the risk expression in (17) and the evaluation $E_0(e^{-Z/\kappa}) = (1 + \frac{2a^2r}{\kappa})^{-\kappa/2}$ (here for $\mu = 0$, $Z \sim \text{Gamma}(\kappa/2, 2a^2r)$) applied for $\kappa = 4$ and $\kappa = 2(c^2+1)$, condition (25), if verified, will indeed imply our result. Finally, we obtain from (17) $(4\pi\sigma_Y^2)^{d/2}\,\Delta_{c^2}(\mu) \;=\; E_{\|\mu\|^2}(g(Z))$, with

$$g(z) = 1 - c^{-d} + 2\left((\frac{2}{c^2+1})^{d/2}\, e^{-\frac{z}{2(c^2+1)}} - e^{-\frac{z}{4}} \right).$$

With (i) $g'(z) \geq 0$ if and only if $z \leq \frac{(d+2)\,\log(\frac{c^2+1}{2})}{\frac{1}{2}-\frac{1}{c^2+1}}$, (ii) $g(0) = 2(\frac{2}{c^2+1})^{d/2} - 1 - c^{-d} < 0$ for all $d, c^2 \geq 1$, and (iii) $\lim_{z\to\infty} g(z) = 1 - c^{-d} > 0$ for all $d \geq 1, c^2 > 1$, we infer that $g(z)$ changes signs exactly once from $-$ to $+$ as z increases on $\mathbb{R}_+$. Consequently, invoking variation diminishing properties (e.g., Brown, Johnstone and MacGibbon, 1981) applicable to the family of $\chi_d^2(\|\mu\|^2)$ distributions of Z, which has an increasing monotone likelihood property in Z with parameter $\|\mu\|^2$, leads to the conclusion that $E_{\|\mu\|^2}(g(Z))$ changes signs at most once as a function of $\|\mu\|^2 \in \mathbb{R}_+$, whence (25) and the desired result. □

Remark 3.5. *For $\mu = 0$, the optimal value c^{*2} of c^2 is available from (20) and the equation $\psi(c^{*2}) = 0$ yielding $c^{*2} = 1 + a^2 r$. Along with (25), we thus infer the lower bound $k_a(d) \geq 1 + a^2 r$.*

[3] With Jensen's inequality and focussing at $\mu = 0$, a necessary condition for this is $E_0(T) \geq 4\log_e(2)$. For $d \geq 3$ and the James-Stein estimator $\hat{\mu}_{JS}(X) = (1 - \frac{(d-2)\sigma_X^2}{X'X})X$, we have $E_0(T) = \frac{1}{\sigma_Y^2}\, E_0(\|\hat{\mu}_{JS}(X) - 0\|^2 = 2r$, so that dominance at $c \to \infty$ is not possible for $\hat{\mu}$ taken to be the James-Stein estimator, or any other estimator dominating $\hat{\mu}_{JS}(X)$ such as its positive part, whenever $r < 4\log_e 2$.

3.3 Improvements over the minimum risk equivariant estimator

As presented in part **(a)** of Theorem 3.1, the point estimation loss $a + bL_{\gamma_0}$ is dual to integrated squared error loss for predictive density estimation with plug-in estimators. Reflected normal loss was introduced by Spiring (1993), namely as an option for a bounded loss. It is also not convex in $\|d-\mu\|$, but strictly bowled shaped in $\|d-\mu\|$. We can thus borrow results applicable to such loss functions. For instance, results from Marchand and Strawderman (2005), or again Kubokawa and Saleh (1994), show that for $d=1$ the Bayes estimator $\hat{\mu}_{\pi_U}(X)$ with respect to the uniform prior either on a compact interval (a,b) or left-bounded interval (a,∞) dominates the MRE estimator X under strictly bowled shaped loss and hence reflected normal loss. Here is a formulation of such an inference.

Theorem 3.3. *For estimating a univariate normal density $Y \sim N(\mu, \sigma_Y^2)$ based on $X \sim N(\mu, \sigma_X^2)$ under integrated squared error loss, and with the restriction $\mu \in [a,b]$ ($\mu \in [a,\infty)$), the estimator $\frac{1}{c\,\sigma_Y}\,\phi(\frac{y-\hat{\mu}_{\pi_U}(X)}{c\,\sigma_Y})$ dominates $\frac{1}{c\,\sigma_Y}\,\phi(\frac{y-X}{c\sigma_Y})$, where $\hat{\mu}_{\pi_U}(X)$ is the Bayes point estimator of μ associated with a uniform prior on $[a,b]$ (on $[a,\infty)$) under reflected normal loss $L_{\gamma_0}(\mu,\hat{\mu})$ with $\gamma_0 = (c^2+1)\sigma_Y^2$.*

Proof. Since $\hat{\mu}_{\pi_U}(X)$ dominates the MRE estimator X as shown by Marchand and Strawderman (2005), the result is a consequence of part **(a)** of Theorem 3.1. □

Another class of applications of part **(a)** of Theorem 3.1 are generated by estimators $\hat{u}(X)$ that dominate X, for $X \sim N_d(\mu, \sigma_X^2 I_d)$ and for loss functions $f(\|d-\mu\|^2)$ with f increasing and concave. Such findings were given by Brandwein and Strawderman (1991, 1981), as well as Brandwein, Ralescu, and Strawderman (1993), and apply for the above reflected normal loss. It is interesting that such a loss arises in our predictive estimation context. The developments that follow make use of similar techniques, but exploit the specific nature of the loss function to obtain a wider class of dominating estimators for $d \geq 3$ of $\hat{q}_{\text{mre}}$. Here is a result of general interest and useful for the developments that follow.

Lemma 3.3. *Let $X \sim N_d(\mu, \sigma_X^2 I_d)$ with known σ_X^2, and consider estimating μ under reflected normal loss $L_\gamma(\mu,\hat{\mu}) = 1 - e^{-\frac{\|\hat{\mu}-\mu\|^2}{2\gamma}}$. Then $\hat{\mu}(X)$ dominates X under L_γ whenever $\hat{\mu}(Z)$ dominates Z for the model $Z \sim N_d(\mu, \frac{\gamma\sigma_X^2}{\gamma+\sigma_X^2}\, I_d)$ under loss $\|\hat{\mu}-\mu\|^2$.*

Proof. Since $-e^{-\frac{\|\hat{\mu}-\mu\|^2}{2\gamma}} = -e^{-\frac{\|x-\mu\|^2}{2\gamma}} \times e^{-\frac{1}{2\gamma}(\|\hat{\mu}-\mu\|^2 - \|x-\mu\|^2)}$, it is seen that

$$L_\gamma(\mu,\hat{\mu}(x)) = 1 - e^{-\frac{\|x-\mu\|^2}{2\gamma}} + e^{-\frac{\|x-\mu\|^2}{2\gamma}} \left(1 - e^{-\frac{1}{2\gamma}(\|\hat{\mu}(x)-\mu\|^2 - \|x-\mu\|^2)}\right),$$

for all $x \in \mathbb{R}^d$. In terms of the risk $R_\gamma(\mu,\hat{\mu}) = E_\mu(L_\gamma(\mu,\hat{\mu}(X))$, we thus have

$$R_\gamma(\mu,\hat{\mu}) = R_\gamma(\mu,X) + E_\mu^X\left[e^{-\frac{\|X-\mu\|^2}{2\gamma}}\left(1 - e^{-\frac{1}{2\gamma}(\|\hat{\mu}(X)-\mu\|^2 - \|X-\mu\|^2)}\right)\right],$$

and

$$\Delta_\gamma(\mu) = R_\gamma(\mu,\hat{\mu}) - R_\gamma(\mu,X) = (\frac{\gamma}{\gamma+\sigma_X^2})^{d/2}\, E_\mu^Z\left(1 - e^{-\frac{1}{2\gamma}(\|\hat{\mu}(Z)-\mu\|^2 - \|Z-\mu\|^2)}\right),$$

where $Z \sim N_d(\mu, \frac{\gamma\sigma_X^2}{\gamma+\sigma_X^2}\, I_d)$. Now, note that $1 - e^{-\eta} < \eta$ for any $\eta \neq 0$, so that

$$\Delta_\gamma(\mu) < (\frac{\gamma}{\gamma+\sigma_X^2})^{d/2}\, E_\mu^Z \left(\|\hat{\mu}(Z) - \mu\|^2 - \|Z - \mu\|^2)\right) , \tag{26}$$

which yields the result. □

Theorem 3.1's duality between the performance **(I)** of plug-in density estimators under integrated squared error loss and the performance **(II)** of the corresponding point estimator under reflected normal loss, coupled with the previous lemma which links the latter's point estimation performance **(II)** with the one under squared error loss **(III)**, lead to the following inadmissibility result and comparisons for our predictive density estimation problem **(I)**.

Theorem 3.4. *For estimating a multivariate normal density $Y \sim N_d(\mu, \sigma_Y^2 I_d)$ based on $X \sim N_d(\mu, \sigma_X^2 I_d)$ under integrated squared error loss, the estimator $\hat{q}_{mre}(\cdot; X) \sim N_d(X, (\sigma_X^2+\sigma_Y^2) I_d)$ is inadmissible for $d \geq 3$, and dominated by any $\hat{q}(\cdot; X) \sim N_d(\hat{\mu}(X), (\sigma_X^2 + \sigma_Y^2) I_d)$, where $\hat{\mu}(Z)$ dominates Z for $Z \sim N_d(\mu, \sigma_Z^2\, I_d)$ under loss $\|\hat{\mu} - \mu\|^2$ and with $\sigma_Z^2 = \frac{(2\sigma_Y^2+\sigma_X^2)\sigma_X^2}{2(\sigma_Y^2+\sigma_X^2)}$.*

Proof. This is a direct consequence of part **(a)** Theorem 3.1 and Lemma 3.3, applied for $c^2 = 1 + \frac{\sigma_X^2}{\sigma_Y^2}$ and $\gamma = (c^2+1)\sigma_Y^2 \; = \; 2\sigma_Y^2 + \sigma_X^2$. □

The above results establishes the inadmissibility of $\hat{q}_{\text{mre}}$ for $d \geq 3$. Along with Stein estimation findings under squared error loss, we can generate explicit dominating plug-in type densities $\hat{p} \sim N_d(\hat{\mu}, (\sigma_X^2 + \sigma_Y^2) I_d)$. Here are some examples.

Example 3.2. *If π is superharmonic prior, the Bayes estimator $\hat{\mu}_\pi(Z)$ dominates Z for $Z \sim N_d(\mu, \sigma_Z^2 I_d)$ for $d \geq 3$ as an estimator of μ under squared error loss (Stein 1981), and the corresponding plug-in density $N_d(\hat{\mu}_\pi(X), (\sigma_X^2+\sigma_Y^2) I_d)$ dominates $\hat{q}_{mre}$ under integrated squared error loss. In terms of the secondary problem of estimating μ under reflected normal loss L_γ, Lemma 3.3 implies the dominance of $\hat{\mu}_\pi(X)$ over X. Wider classes of dominating estimators arise, for instance, by requiring that $\sqrt{m_\pi(z)}$ be superharmonic where $m_\pi(z)$ is the marginal density of Z under π (Fourdrinier, Strawderman, Wells, 1998).*

Example 3.3. *Another class of dominating estimators of $Z \sim N_d(\mu, \sigma_Z^2 I_d)$ for $d \geq 3$, which will arise Section 4 as well, are given by Baranchik type estimators (Baranchik, 1971) $\hat{\mu}_{a,r(\cdot)}(Z) = (1 - a\frac{r(Z'Z)}{Z'Z})\, Z$, such that $r(\cdot)$ is an increasing function, $0 \leq r(\cdot) \leq 1$, $r(\cdot) \neq 0$, and $0 < a \leq 2(d-2)\, \sigma_Z^2$. In view of Theorem 3.4, the corresponding density estimators $N_d(\hat{\mu}_{a,r(\cdot)}(X), (\sigma_X^2 + \sigma_Y^2) I_d)$ dominate $\hat{q}_{mre}$ for $d \geq 3$ under integrated squared error loss.*
As a further illustration, consider the positive part James-Stein estimator $\hat{\mu}_{PJS}(Z) = (1 - \frac{(d-2)\sigma_Z^2}{Z'Z})_+\, Z$, with $a_+ = \max(0, a)$. We proceeded with numerical evaluations, based on Lemma (3.1), of the corresponding predictive density estimator $N_d(\hat{\mu}_{PJS}(X), (\sigma_X^2 + \sigma_Y^2) I_d)$ for various values of $d, \sigma_X^2, \sigma_Y^2$, and in contrast to the constant risk of $\hat{q}_{mre}(\cdot; X) \sim$

$N_d(X, (\sigma_X^2+\sigma_Y^2)I_d)$ given by (15). Figure 1 shows the ratio $\frac{R(\mu,\hat{q}_{PJS})}{R(\mu,\hat{q}_{mre})}$ for $d = 3, 8$, $\sigma_X^2 = 1$, $r = \frac{1}{\sigma_Y^2} = \frac{1}{2}, 1, 2$, as functions of $\lambda = \|\mu\|$. Gains are most important when λ is small (e.g., in the order of 20% for $r = 1$) and dissipate as λ increases, which matches the emblematic behaviour of $\hat{\mu}_{PJS}(Z)$ as an estimator of μ under squared error loss. Furthermore, the gains are amplified when the ratio $r = \frac{\sigma_X^2}{\sigma_Y^2}$ decreases (i.e., larger relative uncertainty in X is better mitigated by the shrinkage procedure). Similar features arise for other dimensions d.

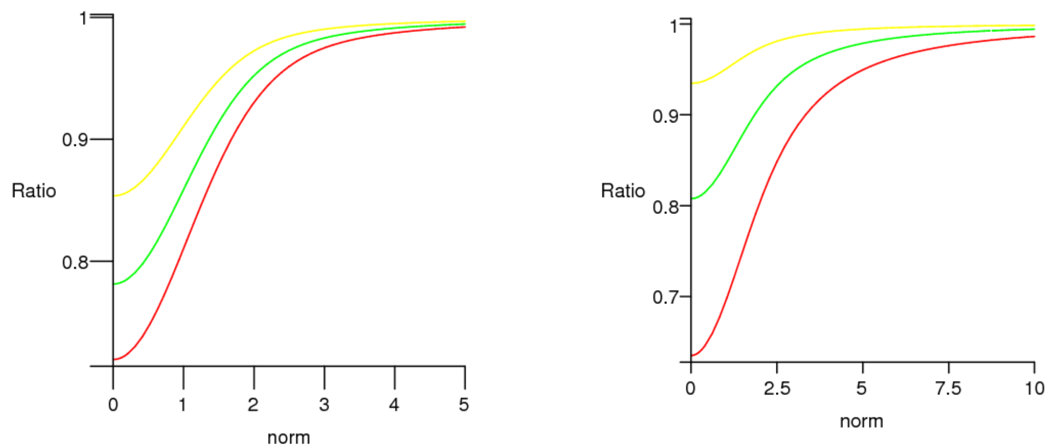

Figure 1: Ratios $\frac{R(\mu,\hat{q}_{PJS})}{R(\mu,\hat{q}_{\rm mre})}$ of risks for $d = 3$(left)$, 8$(right)$, \sigma_X^2 = 1, \sigma_Y^2 = 2, 1, 0.5$, as functions of $\lambda = \|\mu\|$. For fixed λ, ratios increase in σ_Y^2.

Further applications of Theorem 3.4 include the following Hartigan type result for cases where μ is restricted to C, C being a strict subset of $\mathbb{R}^d$ which is convex with a non-empty interior. Such cases include restrictions to balls and to cones such as order constraints $\mu_1 \leq \mu_2 \leq \ldots \leq \mu_d$, tree order constraints $\mu_i \geq \mu_1$ for all i, etc.

Theorem 3.5. *Let $Z \sim N_d(\mu, \sigma_Z^2\, I_d)$ with $\sigma_Z^2 = \frac{(2\sigma_Y^2+\sigma_X^2)\sigma_X^2}{2(\sigma_Y^2+\sigma_X^2)}$, and let $\hat{\mu}_\pi(Z)$ be the Bayes estimator of μ associated with prior density π and loss $\|\hat{\mu} - \mu\|^2$. For estimating the density of $Y \sim N_d(\mu, \sigma_Y^2 I_d)$ based on $X \sim N_d(\mu, \sigma_X^2 I_d)$ under integrated squared error loss, and for the restriction $\mu \in C$ with C a convex subset of $\mathbb{R}^d$ with non-empty interior:*

- **(a)** *the estimator $\hat{q}(\cdot; X) \sim N_d(\hat{\mu}_{\pi_U}(X), (\sigma_X^2+\sigma_Y^2)\, I_d)$ dominates $\hat{q}_{mre}(\cdot; X) \sim N_d(X, (\sigma_X^2+\sigma_Y^2)\, I_d)$ with π_U being the uniform prior on C;*
- **(b)** *for the univariate case with $C = [a, b]$, dominance of $\hat{q}_{mre}(\cdot; X)$ is achieved by any $\hat{q}(\cdot; X) \sim N_d(\hat{\mu}_\pi(X), (\sigma_X^2 + \sigma_Y^2)\, I_d)$ as long as the prior density π is absolutely continuous, symmetric about $(a + b)/2$, and increasing and logconcave on $[\frac{a+b}{2}, b]$.*

Proof. The results are a consequence of part **(a)** of Theorem 3.1 and Lemma 3.3 (with $c^2 = 1+r$) paired with point estimation results of Hartigan (2004) for **(a)**, and Kubokawa (2005) or Marchand and Payandeh Najafabadi (2011) for **(b)**. $\square$

Remark 3.6. *In the context of Theorem 3.4, for cases where* $\|\mu\| \leq m$, *as well as for cases where* $\mu \in C$ *with* C *a convex cone,* $\|\hat{\mu}_{mle}(x) - \mu\|^2$ *is stochastically smaller than* $\|x - \mu\|^2$ *for all* μ *so that* $\hat{\mu}_{mle}(X)$ *dominates* X *as an estimator of* μ *under loss* L_γ, $\gamma > 0$. *Therefore* $\hat{q}_{c^2,\hat{\mu}_{mle}} \sim N_d(\hat{\mu}_{mle}(X), c^2\,\sigma_Y^2\,I_d)$ *dominates* $\hat{q}_{c^2,\hat{\mu}_0} \sim N_d(X, c^2\,\sigma_Y^2\,I_d)$ *for such restricted parameter spaces,* $c^2 > 0$.

4 Extensions to scale mixtures of normals

The developments in this section parallel those of Section 3, but relate to scale mixtures of normals. In Section 4.1, which applies more generally for multivariate location families, we obtain a useful representation for the integrated squared difference (Lemma 4.1) between two densities. This leads to a duality between the frequentist risk of $\hat{q}_{\mathrm{mre}}$, and more generally for density estimators of the form $f(y - \hat{\mu}(X))$, $y \in \mathbb{R}^d$, of a density $q(y - \mu)$, with a risk function for $\hat{\mu}(X)$ (for a loss which we describe) as a point estimator of μ. For scale mixtures of normals, the dual point estimation resulting loss is an non-decreasing and concave function of $\|\hat{\mu} - \mu\|^2$ and, as in Section 3.3, we establish in Section 4.3 the inadmissibility of $\hat{q}_{\mathrm{mre}}$ for $d \geq 3$ (and with risk finiteness conditions), as well as provide dominating estimators. Finally, in Section 4.2, we assess the risk performance of scale expansion estimators of the form $\frac{1}{c^d}\, q(\frac{y-x}{c})$, $c > 1$, in comparison with the plug-in estimator $q(y - x)$, and replicate some of the normal case features with improvements always to be found in this subclass.

4.1 An identity for L_2 distance and general dominance results of plug-in type predictive density estimators

We begin this section with a general identity for integrated squared difference conveniently expressed in terms of convolutions.

Lemma 4.1. *Whenever finite, the integrated squared difference* $\rho_{L_2} = \int_{\mathbb{R}^d} |\,q(t - \mu_1) - f(t-\mu_2)|^2\,dt$ *between densities* $f(t-\mu_1)$ *and* $q(t-\mu_2)$, $\mu_1, \mu_2 \in \mathbb{R}^d$, *is given by* $\rho_{f,q}(\mu_2-\mu_1)$ *with*

$$\rho_{f,q}(s) = q * \bar{q}\,(0) + f * \bar{f}\,(0) - 2q * \bar{f}(s),\ s \in \mathbb{R}^d\,, \tag{27}$$

$\bar{q}(t) = q(-t)$, *and* $\bar{f}(t) = f(-t)$ *for all* $t \in \mathbb{R}^d$.

Proof. In a straightforward manner, we have

$$\begin{aligned}
\rho_{L_2} &= \int_{\mathbb{R}^d} q^2(t-\mu_1)\,dt + \int_{\mathbb{R}^d} f^2(t-\mu_2)\,dt - 2\int_{\mathbb{R}^d} q(t-\mu_1)\,f(t-\mu_2)\,dt \\
&= \int_{\mathbb{R}^d} q(-t)\,\bar{q}(t) + \int_{\mathbb{R}^d} f(-t)\,\bar{f}(t)\,dt \;-\; 2\int_{\mathbb{R}^d} q(\mu_2-\mu_1-t)\,\bar{f}(t)\,dt \\
&= q * \bar{q}\,(0) + f * \bar{f}\,(0) - 2q * \bar{f}(\mu_2-\mu_1) \\
&= \rho_{f,q}(\mu_2-\mu_1). \quad \square
\end{aligned}$$

In our predictive density estimation context, we will be seeking to estimate the density $q(y-\mu)$ under integrated squared error loss, and the above provides the loss associated with the subclass of estimators of the form $f(y-\hat{\mu})$ with f fixed. Comparisons with the MRE estimator, which we carried out for the normal case, are of particular interest. As shown in Example 2.2, such a choice corresponds to $f \equiv q * \bar{p}$ and $\hat{\mu}(x) = x$, with $X \sim p(t-\mu)$ and $\bar{p}(t) = p(-t)$ for all $t \in \mathbb{R}^d$. As a direct consequence of the above Lemma, we have the following.

Corollary 4.1. *For estimating the density $q(y-\mu)$, $y, \mu \in \mathbb{R}^d$, under integrated squared error loss and based on $X \sim p(x-\mu)$,*

- **(a)** *The frequentist risk of the estimator $f(y-\hat{\mu}(X))$ is equal to the frequentist risk of the point estimator $\hat{\mu}(X)$ of μ under loss $\rho_{f,q}(\hat{\mu}-\mu)$;*
- **(b)** *The estimator $f(y-\hat{\mu}_1(X))$ dominates the estimator $f(y-\hat{\mu}_2(X))$ if and only if $\hat{\mu}_1(X)$ dominates $\hat{\mu}_2(X)$ as a point estimator of μ under loss $\rho_{f,q}(\hat{\mu}-\mu)$ or, equivalently, under loss*

$$1 - \frac{2\, q * \bar{f}(\hat{\mu}-\mu)}{q * \bar{q}\,(0) + f * \bar{f}\,(0)}\,; \tag{28}$$

- **(c)** *The estimator $q * \bar{p}(y-\hat{\mu}(X))$ dominates the MRE estimator $q * \bar{p}(y-X)$ if and only if $\hat{\mu}(X)$ dominates X under loss $\rho_{q*\bar{p},q}(\hat{\mu}-\mu)$ or, equivalently, under loss*

$$1 - \frac{2\, q * \bar{q} * p(\hat{\mu}-\mu)}{q * \bar{q}(0) + q * \bar{p} * \bar{q} * p(0)}\,. \tag{29}$$

Proof. Parts **(a)** and **(b)** follow directly from Lemma 4.1. Part **(c)** follows from Proposition 2.1's representation of the MRE estimator and by applying part **(b)** for $f = q * \bar{p}$ and $\bar{f} = \bar{q} * p$. $\square$

Remark 4.1. *With the above results, Corollary 3.1 turns out to be a particular case of Lemma 4.1 by taking q $(= \bar{q}$ by symmetry of $q) \sim N_d(0, \sigma_Y^2 I_d)$, $f(= \bar{f}) \sim N_d(0, c^2\sigma_Y^2 I_d)$, $\mu_1 = \mu$, and $\mu_2 = \hat{\mu}$. Indeed with these values, (27) yields expression (13) with the normal convolutions: $q * \bar{q} \sim N_d(0, 2\sigma_Y^2 I_d)$, $f * \bar{f} \sim N_d(0, 2\, c^2\, \sigma_Y^2 I_d)$, and $q * \bar{f} \sim N_d(0, (1+c^2)\,\sigma_Y^2 I_d)$. Similarly, part* **(a)** *of Theorem 3.1, as well as its dual reflected normal loss $L_{(c^2+1)\sigma_Y^2}$, follow from part* **(b)** *of Corollary 4.1 by setting $f \sim N_d(0, c^2\sigma_Y^2 I_d)$ and $q \sim N_d(0,\ \sigma_Y^2 I_d)$. Namely, the implications of Theorem 3.1 relative to the MRE estimator $\hat{q}_{mre}(\cdot|X)$ (i.e., $c^2 = 1 + r$, $\hat{\mu}_2(X) = X$) are given in part* **(c)** *of Corollary 4.1.*

4.2 Dominating estimators of a plug-in estimator by expanding the scale

We consider here scale mixtures of normals in (1) as described in Definition 2.1:

$$X - \mu \sim SN_d(G)\,,\ \ Y - \mu \sim SN_d(H)\,. \tag{30}$$

As in Section 3.1, we focus on the class of estimators $\hat{q}_c(y;x) = \frac{1}{c^d}\, q(\frac{y-x}{c})$ of the density $q(y-\mu)$, $y \in \mathbb{R}^d$, with $c=1$ corresponding to the plug-in maximum likelihood estimator $q(y-x)$, and choices $c>1$ representing expansions of the known scale associated with the underlying density $q(y-\mu)$. As shown in Theorem 4.1 and Remark 4.2 below, we extend parts **(b)** and **(c)** of Theorem 3.1, applicable to the normal case, which : **(i)** provides the best estimator within the class of $\hat{q}_c$'s; **(ii)** shows the superiority of a subclass of estimators $\hat{q}_c$ over the plug-in $\hat{q}_1$ for a range $(1, c_1)$ (say) of values of c; **(iii)** establishes that $\hat{q}_c$ dominates $\hat{q}_1$ universally for all $c>1$ for sufficiently large dimension d. The next intermediate result follows from Lemma 4.1 and gives the loss incurred by using $\hat{q}_c$ to estimate $q(y-\mu), y \in \mathbb{R}^d$.

Corollary 4.2. *Whenever finite, the integrated squared difference between densities $\hat{q}_c(y;x) = \frac{1}{c^d}\, q(\frac{t-x}{c})$ and $q(t-\mu)$; $t, x, \mu \in \mathbb{R}^d$, $c>0$, is given by*

$$(\frac{1}{c^d}+1)\, q * \bar{q}(0) \;-\; 2q * \bar{h}(x-\mu) \tag{31}$$

where $h(t) = \frac{1}{c^d}\, q(\frac{t}{c})$ and $\bar{h}(t) = h(-t)$ for all t.

Proof. With the evaluation $h * \bar{h}(0) = \frac{1}{c^d}\, q * \bar{q}(0)$, the result follows as Lemma 4.1 by setting $\mu_1 = x, \mu_2 = \mu$, and $f \equiv h$. □

Theorem 4.1. *Let $W_1, W_2 \sim H$, $V_1 \sim G$ be independently distributed. Define $M_c = E[(V_1 + W_1 + c^2\, W_2)^{-d/2}\,]$; $c>1$; $N = E[(W_1+W_2)^{-d/2}\,]$.*[4]

(a) *The risk $R(\mu, \hat{q}_c)$ for estimating a normal scale mixture density $q(t-\mu)$, $t \in \mathbb{R}^d$, with $q \sim SN_d(H)$ under integrated squared error loss and for $X \sim SN_d(G)$, is constant in μ and given by*

$$R(\mu, \hat{q}_c) \;=\; \frac{1}{(2\pi)^{d/2}} \left((1+\frac{1}{c^d})\, N \;-\; 2M_c \right)\,; \tag{32}$$

(b) *The optimal estimator among the class of estimators $\hat{q}_c$ is $\hat{q}_{c^*}$, where c^* is the unique value of $c>1$ such that*

$$2c^{d+2}\, E[\frac{W_2}{(V_1+W_1+c^2\, W_2)^{d/2+1}}] \;=\; N\,; \tag{33}$$

(c) *Estimators $\hat{q}_c$ with $1<c<c_1$ dominate the plug-in $\hat{q}_1$, where $c_1 = \infty$ if $N \geq 2M_1$ and, otherwise, c_1 is the unique value of $c>1$ such that $N(1-\frac{1}{c^d}) = 2(M_1 - M_c)$.*

Proof. (a) The risk of $\hat{q}_c$ is given by

$$R(\mu, \hat{q}_c) \;=\; E_\mu \left[(1+\frac{1}{c^d})\, q * q(0) \;-\; 2\, q * h(X-\mu) \right]\,, \tag{34}$$

by taking the expected value of (31) with respect to $X-\mu \sim SN_d(G)$, and since $\bar{q} \equiv q$ and $\bar{h} \equiv h$ by spherical symmetry of q and thus also h. By making use of Lemma 2.1, we have

[4] The assumptions of Definition 2.1 imply the finiteness of M_c and N.

for the scale mixture of normals density, $q \sim SN_d(H)$, the convolutions $q * q \sim SN_d(F_1)$ and $q * h \sim SN_d(F_c)$, F_c being the cdf of $W_1 + c^2 W_2$ for any c^2. With the above, we obtain

$$q * q(0) \;=\; (2\pi)^{-d/2}\, N\,. \tag{35}$$

Furthermore, we have

$$\begin{aligned}
E_\mu[\, q * h(X-\mu)\,] \;&=\; \int_{\mathbb{R}^d}\int_{\mathbb{R}_+} (2\pi v)^{-d/2} \left(\int_{\mathbb{R}_+} (2\pi t)^{-d/2}\, e^{-\frac{\|x-\mu\|^2}{2t}}\, dF_c(t)\right) e^{-\frac{\|x-\mu\|^2}{2v}}\, dG(v)\, dx \\
&=\; \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} (4\pi^2 vt)^{-d/2} \left(\int_{\mathbb{R}^d} e^{-\frac{\|x\|^2}{2}(\frac{1}{t}+\frac{1}{v})}\, dx\right) dF_c(t) dG(v) \\
&=\; \int_{\mathbb{R}_+}\int_{\mathbb{R}_+} (4\pi^2 vt)^{-d/2} \,(\frac{2\pi\, vt}{v+t})^{d/2}\, dF_c(t) dG(v) \\
&=\; (2\pi)^{-d/2}\, E[(V_1 + W_1 + c^2\, W_2)^{-d/2}] \;=\; (2\pi)^{-d/2}\, M_c\,.
\end{aligned} \tag{36}$$

Finally, the given expression for $R(\mu, \hat{q}_c)$ in (32) follows from (34), (35), and (36).

(b) It is easy to see from (32) that

$$\frac{(2\pi)^{d/2}\, c^{d+1}}{d}\, \frac{\partial}{\partial c} R(\mu, \hat{q}_c) \;=\; -N + 2\, E[\, W_2 (\frac{c^2}{V_1 + W_1 + c^2\, W_2})^{d/2+1}\,]\,.$$

Evaluated at $c = 1$, the above is negative, while it is positive evaluated at $c \to \infty$. Moreover, since the above is increasing as a function of $c \in [1, \infty)$, we have that $\frac{\partial}{\partial c} R(\mu, \hat{q}_c)$ changes signs once from $-$ to $+$ as c increases on $[1, \infty)$ thus establishing the result.

(c) Given that, as a function of c, $R(\mu, \hat{q}_c)$ is strictly decreasing for $1 \leq c < c^*$, and strictly increasing for $c > c^*$, we have indeed $R(\mu, \hat{q}_c) < R(\mu, \hat{q}_1)$ for all $c > 1$ as soon as $\lim_{c\to\infty} R(\mu, \hat{q}_c) \leq R(\mu, \hat{q}_1) \iff 2M_1 \leq N$. Otherwise, we have $R(\mu, \hat{q}_c) < R(\mu, \hat{q}_1)$ for $c < c_1$ with $R(\mu, \hat{q}_{c_1}) = R(\mu, \hat{q}_1) \iff N(1 - \frac{1}{c^d}) = 2(M_1 - M_c)\,.$ □

Remark 4.2. *The above Theorem is presented for fixed d, but there also implications for varying d analogously to part* **(c)** *of Theorem 3.1 established for normal models. Indeed, assuming all the inverse moments associated with H and G exist, which guarantees the finiteness of N and M_1 for all $d \geq 1$, it is inevitable that the interval of values of c such that $\hat{q}_c$ dominates $\hat{q}_1$ is given by $(1, \infty)$ for large enough $d \geq d_0$. This is justified by the fact that if $N \geq 2M_1$ for a given d_0 (which can be shown to exist), i.e.,*

$$E(\frac{1}{(W_1 + W_2)^{d_0/2}}) \geq 2E(\frac{1}{(V_1 + W_1 + W_2)^{d_0/2}})\,,$$

then we must also have

$$E(\frac{1}{(W_1 + W_2)^{(1+d_0)/2}}) \geq 2E(\frac{1}{(V_1 + W_1 + W_2)^{(1+d_0)/2}})\,,$$

telling us that $N \geq 2M_1$ for all $d > d_0$.

Example 4.1. *Theorem 4.1 and Remark 4.2 apply in the normal case with* $P(W_i = \sigma_Y^2) = 1$ *for* $i = 1, 2$ *and* $P(V_1 = \sigma_X^2) = 1$*, and it is readily verified that results in parts* **(b)** *and* **(c)** *of Theorem 3.1 follow.*
As a further illustration, consider situations where W_1, W_2, V_1 *share the same distribution with* $P(r_1 \leq W_1 \leq r_2) = 1$ *for some* $0 < r_1 \leq r_2 < \infty$*. By setting* $T = V_1 + W_1 =^d W_1 + W_2$ *and* $D = W_2/T$*, equation (33) may be expressed as*

$$c^{d+2}\, E^T \left(T^{-\frac{d}{2}} E^{D|T} \left(\frac{2D}{(1+c^2 D)^{d/2+1}} \right) \right) = E^T \left(T^{-\frac{d}{2}} \right) ,$$

so that the condition

$$C(t) \;=\; c_0^{d+2}\, E^{D|T=t} \left(\frac{2D}{(1+c_0^2 D)^{d/2+1}} \right) \leq 1 , \tag{37}$$

for all $t \in [2r_1, 2r_2]$*, and for some* c_0*, implies* $c^* \geq c_0$*. Using the covariance inequality* $Cov(f_1(D), f_2(D)) \leq 0$ *for increasing* f_1 *and decreasing* f_2*, as well as the property* $E(D|T) = 1/2$ *which is a consequence of the iid assumption on* W_1, W_2, V_1*, we obtain*

$$C(t) \leq c_0^{d+2}\, E^{D|T} \left(\frac{1}{(1+c_0^2 D)^{d/2+1}} \right) .$$

Now, we have with the bounded support assumption and by setting $\beta = \frac{r_1}{r_1+r_2}$*,* $P(D \geq \beta | T = t) = 1$ *and* $C(t) \leq (\frac{c_0^2}{1+c_0^2\beta})^{d/2+1}$ *for all* t*. Finally, setting this upper bound on* $C(t)$ *equal to* 1*, we obtain the lower bound*

$$c^{*2} \geq 1 + \frac{r_1}{r_2} .$$

Turning now to a value d_0 *such that all* $\hat{q}_c$ *with* $c > 1$ *dominate the plug-in* $\hat{q}_1$ *for all* $d \geq d_0$*, Theorem 4.1's condition* $N \geq 2M_1$ *may be written as*

$$N \geq 2M_1 \Longleftrightarrow E(T^{-\frac{d}{2}}) \geq 2E^T \left(T^{-\frac{d}{2}}\; E^{D|T} (1+D)^{-\frac{d}{2}} \right) ,$$

which becomes satisfied as soon as $E^{D|T=t}(1+D)^{-\frac{d}{2}} \leq \frac{1}{2}$ *for all* t*. Finally, with* $P(D \geq \beta | T = t) = 1$*, we conclude that all* $\hat{q}_c$ *with* $c > 1$ *dominate the plug-in* $\hat{q}_1$ *for all* $d \geq d_0 = \frac{\log 4}{\log(1+\beta)}$ *.*

4.3 Inadmissibility of the MRE density estimator for $d \geq 3$ and dominating estimators

Despite the fact that Lemma 4.1 and parts **(a)** and **(b)** of Corollary 4.1 apply for general densities q, f, p, we will focus here on applications of part **(c)** of Corollary 4.1 for scale mixtures of normals. As in Section 3.3, we exploit the property that the dual loss in (28) is an increasing and concave function of $\|\hat{\mu} - \mu\|^2$, to derive dominating estimators of X.

The first part of the following result is an adaptation of part **(c)** of Corollary 4.1 for scale mixtures of normals and for comparing estimators with $\hat{q}_{\mathrm{mre}}$, the middle part establishes that point estimation dominance results with squared-error loss under an associated scale mixture of normals model generates dominating estimators of $\hat{q}_{\mathrm{mre}}$, and the last part capitalizes on an existing result (Strawderman, 1974) for scale mixtures of normals and leads to an inadmissibility result for $\hat{q}_{\mathrm{mre}}$ and $d \geq 3$.

Theorem 4.2. *Consider estimating a scale mixture of normals density $q(y-\mu)$, $y \in \mathbb{R}^d$, of $Y \sim SN_d(H)$ under integrated squared error loss, and based on $X \sim SN_d(G)$ having density $p(x-\mu)$. Let $W_1, W_2 \sim H$, $V_1 \sim G$ be independently distributed, and let F and J be the cdfs of $V_1 + W_1$ and $V_1 + W_1 + W_2$ respectively.*

(a) *The estimator $q * p(y - \hat{\mu}(X))$ dominates the MRE estimator $q * p(y - X)$, with $q * p \sim SN_d(F)$, if and only if $\hat{\mu}(X)$ dominates X under loss*

$$f(\|\hat{\mu}-\mu\|^2) \;=\; K - \int_{\mathbb{R}_+} (2\pi\, t)^{-d/2} e^{-\frac{\|\hat{\mu}-\mu\|^2}{2t}} \; dJ(t)\,, \tag{38}$$

K being a constant.

(b) *The estimator $q * p(y - \hat{\mu}(X))$ dominates the MRE estimator $q * p(y - X)$, with $q * p \sim SN_d(F)$, whenever $\hat{\mu}(X')$ dominates X' under squared-error loss $\|\hat{\mu}-\mu\|^2$ and for $X' - \mu \sim SN_d(F_Z)$, F_Z being the cdf of*

$$Z =^d \frac{Z_1 Z_2}{Z_1 + Z_2}\,,\;\; \text{with } (Z_1, Z_2) \sim d\tau(z_1, z_2) \propto \frac{1}{z_2\,(z_1+z_2)^{d/2}}\, dG(z_1)\, dJ(z_2)\,. \tag{39}$$

(c) *Assuming $E(Z)$ and $E(Z^{-1})$ exist, $\hat{q}_{\mathrm{mre}}$ is inadmissible for $d \geq 3$ and dominated by $q * p(y - \hat{\mu}_{a,r(\cdot)}(X))$ with $q * p \sim SN_d(F)$, and with a Baranchik type estimator $\hat{\mu}_{a,r(\cdot)}(X) = (1 - a\frac{r(X'X)}{X'X})\, X$ such that $r(\cdot)$ is an increasing function, $\frac{r(t)}{t}$ decreases in t, $0 \leq r(\cdot) \leq 1$, $r(\cdot) \neq 0$, and $0 < a \leq \frac{2(d-2)}{E(Z^{-1})}$.*

Proof. Part **(a)** follows from part **(a)** of Corollary 4.1 and Lemma 2.1's convolution properties for scale mixtures of normals. For part **(b)**, we seek a condition that suffices for the difference in risks $\Delta(\hat{\mu}, \mu) = E_\mu\,[\, f(\|\hat{\mu}(X) - \mu\|^2) \;-\; f(\|X-\mu\|^2)\,]$ to be less than 0, where f is given in (38). Since f is strictly concave, the inequality $f(s) - f(t) < f'(t)(s-t)$, for $s \neq t$ and for such f's, implies for the difference in losses that

$$f(\|\hat{\mu}(x)-\mu\|^2) \;-\; f(\|x-\mu\|^2) < f'(\|x-\mu\|^2)\,\left(\|\hat{\mu}(x)-\mu\|^2 - \|x-\mu\|^2\right)\,, \tag{40}$$

for all $x, \mu \in \mathbb{R}^d$ such that $x \neq \hat{\mu}(x)$. Now, using the above, part **(b)** follows since

$$\begin{aligned}
\Delta(\hat{\mu},\mu) \;&<\; \frac{1}{2} E_\mu^X \left[\left(\|\hat{\mu}(X)-\mu\|^2 - \|X-\mu\|^2\right) \int_0^\infty (2\pi t)^{-d/2}\, e^{-\frac{\|X-\mu\|^2}{2t}}\, \frac{dJ(t)}{t} \right] \\
&=\; \frac{1}{2} \int_{\mathbb{R}^d} \left(\|\hat{\mu}(x)-\mu\|^2 - \|x-\mu\|^2\right) \int_0^\infty \int_0^\infty (4\pi^2\, st)^{-\frac{d}{2}}\, e^{-\frac{\|x-\mu\|^2}{\frac{2st}{s+t}}}\, dG(s)\, \frac{dJ(t)}{t}\, dx \\
&\propto\; \int_{\mathbb{R}^d} \left(\|\hat{\mu}(x)-\mu\|^2 - \|x-\mu\|^2\right) \int_0^\infty \int_0^\infty (\frac{2\pi\, st}{s+t})^{-\frac{d}{2}}\, e^{-\frac{\|x-\mu\|^2}{\frac{2st}{s+t}}}\, d\tau(s,t)\, dx \\
&\propto\; \int_{\mathbb{R}^d} \left(\|\hat{\mu}(x)-\mu\|^2 - \|x-\mu\|^2\right) \int_0^\infty (2\pi z)^{-\frac{d}{2}}\, e^{-\frac{\|x-\mu\|^2}{2z}}\, dF_Z(z)\; dx\,.
\end{aligned}$$

Finally, part **(c)** follows from pairing Theorem 2.1 of Strawderman (1974) with part **(b)** above, for the model $X' - \mu \sim SN_d(F_Z)$, with the finiteness conditions $E_0\|X'X\| = dE(Z) < \infty$ and $E_0\|X'X\|^{-1} = \frac{1}{d-2}\,E(Z^{-1}) < \infty$, and with the upper bound on the constant a given by $\frac{2}{E_0\|X'X\|^{-1}}$. □

Remark 4.3. *The dual loss in (38) may be labelled as reflected scale mixture of normals analogously to reflected normal loss. For the normal case with degenerate W_1, W_2, V_1 at σ_Y^2 and σ_X^2 respectively, part* **(a)** *of Theorem 4.2 applied to the MRE estimator reduces to Corollary 3.1, while part* **(b)** *reduces to Theorem 3.4 with Z in (39) degenerate at $\frac{\sigma_X^2(\sigma_X^2+2\sigma_Y^2)}{2(\sigma_X^2+\sigma_Y^2)}$.*

We pursue with some examples of applications of Theorem 4.2.

Example 4.2. *(Cases where both H and G are Gamma cdf's) We illustrate some of the features of Theorem 4.2 for situations where $W_1, W_2 \sim H \sim Gamma(\alpha_1, 1)$, $V_1 \sim G \sim Gamma(\alpha_2, 1)$, with $\alpha_1 > d/2$ for $i = 1, 2$ which guarantees that $E(V_1^{-d/2}) < \infty$ and $E(W_1^{-d/2}) < \infty$ (see Definition 2.1). We have $V_1 + W_1 \sim F \sim Gamma(\alpha_1 + \alpha_2, 1)$ and $V_1 + W_1 + W_2 \sim J \sim Gamma(\alpha_1 + 2\alpha_2, 1)$. Part* **(d)** *of Example 2.3 indicates that*

$$\hat{q}_{mre}(y; x) \;=\; q * p(y - x), \;\; with \;\; q * p \sim SN_d(F)\,.$$

Theorem 4.2 tells us that $\hat{q}_{mre}(y; X)$ is inadmissible as an estimator of $q(y - \mu)$, $y \in \mathbb{R}^d$ for $d \geq 3$, and dominated by any $q*p(y - \hat{\mu}(X))$, where $\hat{\mu}(X')$ dominates X' under squared error loss and for $X' - \mu \sim SN_d(F_Z)$ as given in (39). The joint density of (Z_1, Z_2) in (39) becomes

$$d\tau(z_1, z_2) \;=\; k\frac{z_1^{\alpha_1-1}\; z_2^{\alpha_1+2\alpha_2-2}\; e^{-(z_1+z_2)}}{(z_1+z_2)^{d/2}}\; \mathbb{I}_{\mathbb{R}_+}(z_1)\,\mathbb{I}_{\mathbb{R}_+}(z_2)\,,$$

with $k = \frac{\Gamma(2\alpha_1+2\alpha_2-1)}{\Gamma(\alpha_1)\,\Gamma(\alpha_1+2\alpha_2-1)\,\Gamma(2\alpha_1+2\alpha_2-1-d/2)}$, and where we have used the identity

$$\int_{\mathbb{R}_+}\int_{\mathbb{R}_+} \frac{z_1^{a-1}\; z_2^{b-1}}{(z_1+z_2)^c}\, e^{-(z_1+z_2)}\, dz_1 dz_2 \;=\; \frac{\Gamma(a)\,\Gamma(b)\,\Gamma(a+b-c)}{\Gamma(a+b)}\,,$$

for $a, b > 0$, $a + b > c$. Finally, using again the above identity, calculations yield

$$E_\tau(Z^{-1}) = E_\tau(Z_1{}^{-1}) + E_\tau(Z_2{}^{-1}) \;=\; \frac{(\alpha_1+\alpha_2-1)(2\alpha_1+2\alpha_2-3)}{(\alpha_1+\alpha_2-1-d/4)\,(\alpha_1-1)\,(\alpha_1+2\alpha_2-2)}\,,$$

so that part **(c)**'s subclass of dominating Baranchik predictive density estimators is explicitly determined with $0 < a \leq \frac{2(d-2)}{E_\tau(Z^{-1})}$ and the above $E_\tau(Z^{-1})$.

Example 4.3. *(Cases where the mixing distribution G is lower bounded) Consider situations where either $X - \mu \sim N_d(0, \sigma_X^2 I_d)$ or, more generally, the scale parameter distribution for X is bounded below by some known $a_X > 0$ (i.e., $G^-(a_X) = 0$ where G^- is the left-hand limit of G). With such an assumption, without any additional knowledge on G, one can obtain an upper bound for Theorem 4.2's $E(Z^{-1})$ and, hence, a lower* *bound for part* **(c)***'s upper limit* $\frac{2(d-2)}{E(Z^{-1})}$*. Indeed, the lower bound assumption implies that* $P(Z_1 \geq a_X) = 1$*,* $P(Z_2 \geq a_X) = 1$*,* $E(Z^{-1}) \leq \frac{2}{a_X}$*, and* $0 < a < (d-2)a_X$ *for Theorem 4.2's Baranchik-type estimators* $\hat{\mu}_{a,r(\cdot)}$*. Similarly, if the mixing variance distribution* H *for* Y *is bounded below by some* $a_Y > 0$*, the above bounds become* $E(Z^{-1}) \leq \frac{1}{a_X} + \frac{1}{a_X + 2a_Y}$*, and* $0 < a \leq \frac{(d-2)a_X(a_X+2a_Y)}{a_X+a_Y}$*; with the degenerate case bringing us back to Example 3.3.*

5 Concluding Remarks

In summary, the findings of this paper provide fundamental identities and results for assessing the efficiency in terms of frequentist risk of predictive density estimators of multivariate observables for integrated squared error loss. For multivariate normal models, we have established a connection between the average integrated squared error loss of plug-in type estimators and point estimation risk under reflected normal loss. Paired with Stein estimation techniques and results for estimating a multivariate normal mean under loss which is a concave function of the squared error $\|\hat{\mu} - \mu\|^2$, we establish the inadmissibility of the minimum risk equivariant (MRE) density estimator and obtain dominating predictive density estimators for three dimensions or more. The duality is further exploited to obtain improvements of the benchmark MRE density estimator in the presence of restrictions on the underlying mean parameter. We have also analyzed the performance of scale expansion plug-in density estimators $N_d(\hat{\mu}(X), c^2\sigma_Y^2 I_d)$ with varying c^2, obtaining notably instances (i.e., large enough dimension d and $\hat{\mu}(x) = ax$ with $0 < a \leq 1$) where all scale expansions $c^2 > 1$ improve uniformly on $c^2 = 1$.

For scale mixtures of multivariate normal observables, we have obtained analogous developments with regards to the MRE density estimator by making use of a general integrated squared difference identity, including its inadmissibility and the determination of explicit improvements, in general for three of more dimensions. As well, we obtain improvements on the plug-in maximum likelihood estimator by scale expansion.

One notable drawback of improved plug-in estimators, or improved variance-inflated variants, is that they are generally not (generalized) Bayes estimators. For Kullback-Leibler loss and normal models, there exists an elegant duality relationship between Bayesian predictive densities and Bayesian point estimators under the same priors (e.g., George, Liang, Xu, 2006). Establishing similar connections for integrated squared error loss and normal models other than for plug-in density estimators, remains an open and challenging problem. However, our findings, and specifically the duality for plug-in estimators, apply to scale mixtures of normals for which Bayesian dual relationships are, to our knowledge, unavailable.

6 Appendix

A.1. Minimax estimator and least favourable sequence in the normal case

We provide here for normal case (3) a direct approach to obtain a least favorable sequence of priors and show that the best equivariant estimator $\hat{q}_{\rm mre}(\cdot|X) \sim N_d(X, (\sigma_X^2+\sigma_Y^2)\, I_d)$ is minimax under integrated squared error loss. We proceed in a familiar way showing that $\hat{q}_{\rm mre}(\cdot|X)$ is extended Bayes with constant risk. We make use of (15), which established that the constant risk of $\hat{q}_{\rm mre}(\cdot|X)$ is given by $R_0 = (4\pi\sigma_Y^2)^{-d/2} - (4\pi(\sigma_X^2+\sigma_Y^2))^{-d/2}$. Consider the sequence of priors $\pi_m \sim N_d(0, mI_d)$; $m = 1, 2, \ldots$; as in Example 2.1 with $\theta = 0, \tau^2 = m$, and with corresponding Bayes estimators $\hat{q}_{\pi_m}(\cdot|X) \sim N_d(\hat{\mu}_{\pi_m}(X), \sigma_{1,m}^2\, I_d)$, $\hat{\mu}_{\pi_m}(x) = \frac{mx}{m+\sigma_X^2}$, and $\sigma_{1,m}^2 = \frac{m\sigma_X^2}{m+\sigma_X^2} + \sigma_Y^2$. The posterior loss $\int_{\mathbb{R}^d} (\hat{q}_{\pi_m}(y|x) - q(y-\mu)\,)^2\, dy$ is obtained from (7) with $\sigma_1^2 = \sigma_{1,m}^2$, $\sigma_2^2 = \sigma_Y^2$, $\mu_1 = \hat{\mu}_{\pi_m}(x)$, $\mu_2 = \mu$, and given by

$$k_m(\mu, x) = \frac{1}{(4\pi\sigma_{1,m}^2)^{d/2}} + \frac{1}{(4\pi\sigma_Y^2)^{d/2}} - \frac{2}{(\sigma_{1,m}^2+\sigma_Y^2)^{d/2}}\, \phi\left(\frac{\hat{\mu}_{\pi_m}(x) - \mu}{\sqrt{\sigma_{1,m}^2+\sigma_Y^2}}\right).$$

From this and by making use of (6), the expected posterior loss is evaluated as

$$\begin{aligned}\int_{\mathbb{R}^d} k_m(\mu)\, \pi(\mu|x)\, d\mu &= \frac{1}{(4\pi\sigma_{1,m}^2)^{d/2}} + \frac{1}{(4\pi\sigma_Y^2)^{d/2}} \\ &- \frac{2}{(\sigma_{1,m}^4 - \sigma_Y^4)^{d/2}} \int_{\mathbb{R}^d} \phi\left(\frac{\hat{\mu}_{\pi_m}(x) - \mu}{\sqrt{\sigma_{1,m}^2+\sigma_Y^2}}\right) \phi\left(\frac{\hat{\mu}_{\pi_m}(x) - \mu}{\sqrt{\sigma_{1,m}^2-\sigma_Y^2}}\right) dy \\ &= \frac{1}{(4\pi\sigma_{1,m}^2)^{d/2}} + \frac{1}{(4\pi\sigma_Y^2)^{d/2}} - \frac{2}{(4\pi\sigma_{1,m}^2)^{d/2}}. \qquad (41)\end{aligned}$$

Observe that the expected posterior loss, given in (41), is independent of x and thus matches the Bayes risk r_{π_m}. Finally, since $\sigma_{1,m}^2 \to \sigma_X^2 + \sigma_Y^2$ when $m \to \infty$, we have $\lim_{m\to\infty} r_{\pi_m} = R_0 = R(\mu, \hat{q}_{\rm mre})$, which implies that the estimator $\hat{q}_{\rm mre}$ is indeed minimax (and that the sequence π_m is least favorable).

A.2. Proof of the minimaxity in Proposition 2.1.

We proceed as in Girshick and Savage (1951). For $\mu = (\mu_1, \ldots, \mu_d)'$, let $A_k = \{\mu|\ |\mu_i| < k/2, i = 1, \ldots, d\}$ for $k = 1, 2, \ldots$, and consider the sequence of prior distributions given by

$$\pi_k(\mu) = \begin{cases} k^{-d} & \text{if } \mu \in A_k \\ 0 & \text{otherwise,} \end{cases}$$

which yields the Bayes estimators

$$\hat{q}_k^\pi(y|x) = \int_{A_k} q(y-a)\, p(x-a)\, da \Big/ \int_{A_k} p(x-a)\, da$$

with the Bayes risk function

$$r_k(\pi_k, \hat{q}_k^\pi) = \frac{1}{k^d} \int_{A_k} \int \int \{q(y-\mu) - \hat{q}_k^\pi(y|x)\}^2\, dy\, p(x-\mu)\, dx\, d\mu.$$

Since $r_k(\pi_k, \hat{q}_k^\pi) \le r_k(\pi_k, \hat{q}\mathrm{mre}) = R(\mu, \hat{q}\mathrm{mre}) \equiv R_0$, it is sufficient to show that $\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi) \ge R_0$. Making the transformations $z = x - \mu$ $(dz = dx)$, $t = a - \mu$ $(dt = da)$ and $\xi_i = \mu_i/k$ $(d\xi_i = d\mu_i/k)$ gives that

$$\begin{aligned} r_k(\pi_k, \hat{q}_k^\pi) &= \int_{|\xi_i|<1/2, i=1,\ldots,d} \int\int \{q(y) - \hat{q}_k^{\pi*}(y|x)\}^2 \, dy \, p(x) dx \, d\mu \\ &\ge \int_{|\xi_i|<(1-\epsilon)/2, i=1,\ldots,d} \int\int \{q(y) - \hat{q}_k^{\pi*}(y|x)\}^2 \, dy \, p(x) \, dx \, d\mu, \end{aligned}$$

for any $\epsilon > 0$, where

$$\hat{q}_k^{\pi*}(y|x) = \int_{t+k\xi\in A_k} q(y-t)\, p(x-t)\, dt \Big/ \int_{t+k\xi\in A_k} p(x-t)\, dt.$$

For $|\xi_i| < (1-\epsilon)/2$, it is seen that $\{t + k\xi \in A_k\} \supset \{-k\epsilon/2 < t_i < k\epsilon/2, i = 1, \ldots, d\}$, which implies that $\hat{q}_k^{\pi*}(y|x) \to \hat{q}_{mre}(y|x)$ as $k \to \infty$. Using Fatou's lemma, one gets

$$\begin{aligned} \liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi) &\ge \liminf_{k\to\infty} \int_{|\xi_i|<(1-\epsilon)/2, i=1,\ldots,d} \int\int \{q(y) - \hat{q}_k^{\pi*}(y|x)\}^2 \, dy \, p(x) \, dx \, d\mu, \\ &\ge \int_{|\xi_i|<(1-\epsilon)/2, i=1,\ldots,d} \int\int \liminf_{k\to\infty} \{q(y) - \hat{q}_k^{\pi*}(y|x)\}^2 \, dy \, p(x) \, dx \, d\mu, \\ &= (1-\epsilon)^d R(\mu, \hat{q}_{mre}(y|x)) = (1-\epsilon)^d R_0 \,. \end{aligned}$$

From the arbitrariness of $\epsilon > 0$, it follows that $\liminf_{k\to\infty} r_k(\pi_k, \hat{q}_k^\pi) \ge R_0$, which proves the minimaxity of the MRE predictor. □

Acknowledgements

The authors are grateful to the Associate Editor and a diligent referee for their valuable comments and helpful suggestions. Tatsuya Kubokawa's research is supported in part by Grant-in-Aid for Scientific Research (23243039 and 26330036) from the Japan Society for the Promotion of Science, Eric Marchand's research is supported in part by the Natural Sciences and Engineering Research Council of Canada, and William Strawderman's research is partially supported by a grant from the Simons Foundation (#209035). We are grateful to the Center for International Research on the Japanese Economy which provided financial support for a 2012 visit by Marchand to the University of Tokyo.

References

[1] Aitchison, J. (1975). Goodness of prediction fit. *Biometrika*, **62**, 547-554.

[2] Andrews, D.R. and Mallows, C.L. (1974). Scale mixtures of normal distributions. *Journal of the Royal Statistical Society B*, **36**, 99-102.

[3] Baranchik, A.J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution. *Annals of Mathematical Statistics*, **41**, 642-645.

[4] Berg, C. and Vignat, C. (2010). On the density of the sum of two independent Student t-random vectors. *Statistics & Probability Letters*, **80**, 1043-1055.

[5] Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities. *Annals of Statistics*, **3**, 1318-1328.

[6] Brandwein, A.C., Ralescu, S. and Strawderman, W.E. (1993). Shrinkage estimators of the location parameter for certain spherically symmetric distributions. *Annals of the Institute of Statistical Mathematics*, **45**, 551-565.

[7] Brandwein, A.C and Strawderman, W.E. (1991). Generalization of James-Stein estimators under spherical symmetry. *Annals of Statistics*, **19**, 639-650.

[8] Brandwein, A.C and Strawderman, W.E. (1981). Minimax estimation of location parameters for spherically symmetric distributions with concave loss. *Annals of Statistics*, **8**, 279-284.

[9] Brown, L.D., George, E.I., and Xu, X. (2008). Admissible predictive density estimation. *Annals of Statistics*, **36**, 1156-1170.

[10] Brown, L.D., Johnstone, I.M., and MacGibbon, K.B. (1981). Variation diminishing transformations: A direct approach to total positivity and its statistical applications. *Journal of the American Statistical Association*, **76**, 824-832.

[11] Eaton, M.L. (1989). *Group Invariance Applications in Statistics*. Regional conference series in Probability and Statistics, vol. 1, Institute of Mathematical Statistics and the American Statistical Association.

[12] Feller, W. (1966). *An Introduction to Probability Theory and its Application*, **2**, second edition, Wiley, New York.

[13] Fourdrinier, D., Marchand, É., Righi, A. and Strawderman, W.E. (2011). On improved predictive density estimation with parametric constraints. *Electronic Journal of Statistics*, **5**, 172-191.

[14] Fourdrinier, D., Strawderman, W.E., and Wells, M. (1998). On the construction of Bayes minimax estimators. *Annals of Statistics*, **26**, 660-671.

[15] George, E. I., Liang, F. and Xu, X. (2006). Improved minimax predictive densities under Kullback-Leibler loss. *Annals of Statistics*, **34**, 78-91.

[16] Girshick, M.A. and Savage, L.J. (1951). Bayes and minimax estimates for quadratic loss functions. In *Proc. Second Berkeley Symp. Math. Statist. Probab.*, **1**, 53-74. University of California Press, Berkeley.

[17] Hartigan, J. (2004). Uniform priors on convex sets improve risk. *Statistics & Probability Letters*, **67**, 285-288.

[18] Kano, Y. (1994). Consistency property of elliptical probability density functions. *Journal of Multivariate Analysis*, **51**, 139-147.

[19] Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. *Annals of Mathematical Statistics*, **28**, 573-601.

[20] Komaki, F. (2001). A shrinkage predictive distribution for multivariate normal observables. *Biometrika*, **88**, 859-864.

[21] Kubokawa, T., Marchand, É., Strawderman, W.E. and Turcotte, J-P. (2013). Minimaxity in predictive density estimation with parametric constraints. *Journal of Multivariate Analysis*, **116**, 382-397.

[22] Kubokawa, T., Marchand, É. and Strawderman, W.E. (2015). On improved shrinkage estimators under concave loss. *Statistics & Probability Letters*, **96**, 241-246.

[23] Kubokawa, T. and Saleh, A.K.MD.E. (1998). Estimation of location and scale parameters under order restrictions. *Journal of Statistical Research*, **28**, 41-51.

[24] Lawless, J.F. and Fredette, M. (2005). Frequentist prediction intervals and predictive distributions. *Biometrika*, **92**, 529-542.

[25] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation.* Springer-Verlag, 2nd edition.

[26] Liang, F. and Barron, A. (2004). Exact minimax strategies for predictive density estimation, data compression, and model selection. *IEEE Trans. Inform. Theory*, **50**, 2708-2726.

[27] Marchand, É., and Payandeh Najafabadi, A.T. (2011). Bayesian improvements of a MRE estimator of a bounded location parameter. *Electronic Journal of Statistics*, **5**, 1495-1502.

[28] Marchand, É., and Strawderman, W. E. (2005). On improving on the minimum risk equivariant estimator of a location parameter which is constrained to an interval or a half-interval. *Annals of the Institute of Statistical Mathematics*, **57**, 129-143.

[29] Marchand, É. and Strawderman, W.E. (2004). Estimation in restricted parameter spaces: A review. *Festschrift for Herman Rubin*, IMS Lecture Notes-Monograph Series, **45**, 21-44.

[30] Murray, G.D. (1977). A note on the estimation of probability density functions. *Biometrika*, **64**, 150-152.

[31] Nadarajah, S. (2003). The Kotz-type distribution with applications. *Statistics*, **37**, 341-358.

[32] Nason, G.P. (2006). On the sum of t and Gaussian random variables. *Statistics & Probability Letters*, **76**, 1280-1055.

[33] Shao, J. (1999). *Mathematical statistics.* Springer Texts in Statistics. Springer-Verlag, New York.

[34] Spiring, F.A. (1993). The reflected normal loss function. *Canadian Journal of Statistics*, **31**, 321-330.

[35] Stein, C. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, **9**, 1135-1151.

[36] Strawderman, W.E. (1974). Minimax estimation of location parameters for certain spherically symmetric distributions. *Journal of Multivariate Analysis*, **4**, 255-264.